DOCUMENT RESUME

ED 144 958 TM 006 391

AUTHOR	Wearne, Diana Catherine
TITLE	Development of a Test of Mathematical Problem Solving Which Yields A Comprehension, Application, and Problem Solving Score. Technical Report No. 407.
INSTITUTION	Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	WRDCCL-TR-407
PUB DATE	Dec 76
CONTRACT	NE-C-00-3-0065
NOTE	174p.; Report from the Project on Conditions of Learning and Instructional Strategies ; Parts of appendix may be marginally legible
EDRS PRICE	MF-$0.83 HC-$8.69 Plus Postage.
DESCRIPTORS	Complexity Level; Elementary Education; Elementary School Students; Grade 4; *Item Analysis; Mathematical Applications; Mathematical Concepts; *Mathematics; *Problem Solving; *Test Construction; *Test Items; Test Reliability; Test Validity
IDENTIFIERS	*Problem Solving Test (Wis Cognitive Learning)

ABSTRACT

A test of problem solving behavior which provides information about the mastery of the prerequisites of the problems has been developed for fourth grade children. Each problem solving question is preceded by two other questions which assess the child's understanding of the information contained in the actual question, and an application question which measures the child's knowledge of an underlying concept of the problem solving question. The study is concerned with determining whether asking multiple questions about the same unit of information affects the response to those questions, and with measuring the extent to which the questions preceding the problem solving question assess prerequisites of it. The test was administered to 600 fourth grade students in Wisconsin and the resulting data are presented. Various types of item analysis are described; reliability and content validity of the test are reported. A copy of The Problem Solving Test is appended. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

Technical Report No. 407

Development of a Test of Mathematical Problem Solving Which Yields a Comprehension, Application, and Problem Solving Score

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: TM | SE
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Regents of Univ. of Wisconsin
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

ED144958

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

December 1976

Wisconsin Research and
Development Center for
Cognitive Learning

TM006 391

Technical Report No. 407

DEVELOPMENT OF A TEST OF MATHEMATICAL PROBLEM SOLVING WHICH YIELDS A COMPREHENSION, APPLICATION, AND PROBLEM SOLVING SCORE

by

Diana Catherine Wearne

Report from the Project on
Conditions of Learning and Instructional Strategies

Thomas A. Romberg
Faculty Associate

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

December 1976

This Technical Report is a doctoral dissertation reporting research supported by the Wisconsin Research and Development Center for Cognitive Learning. Since it has been approved by a University Examining Committee, it has not been reviewed by the Center. It is published by the Center as a record of some of the Center's activities and as a service to the student. The bound original is in the University of Wisconsin Memorial Library.

Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the National Institute of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by that agency should be inferred.

Center Contract No. NE-C-00-3-0065

WISCONSIN RESEARCH AND DEVELOPMENT CENTER FOR COGNITIVE LEARNING

MISSION

The mission of the Wisconsin Research and Development Center for Cognitive Learning is to help learners develop as rapidly and effectively as possible their potential as human beings and as contributing members of society. The R&D Center is striving to fulfill this goal by

- conducting research to discover more about how children learn
- developing improved instructional strategies, processes and materials for school administrators, teachers, and children, and
- offering assistance to educators and citizens which will help transfer the outcomes of research and development into practice

PROGRAM

The activities of the Wisconsin R&D Center are organized around one unifying theme, Individually Guided Education.

FUNDING

The Wisconsin R&D Center is supported with funds from the National Institute of Education; the Bureau of Education for the Handicapped, U.S. Office of Education; and the University of Wisconsin.

ACKNOWLEDGMENTS

The study described in this paper would not have been possible without the assistance of the 600 fourth grade children who participated in the study, Kurt Ruthmansdorfer, the programmer at MACC, and the financial assistance of the U.S. Government and the State of Wisconsin.

I would like to thank my major professor, Thomas A. Romberg, for encouraging my interests in assessment questions and for suggesting avenues of study I otherwise might not have pursued. The other two members of my committee, Professors John G. Harvey and J. Fred Weaver, also deserve thanks - Professor Harvey for introducing me to the AMI project thereby securing for me not only lasting friendships among the DMP staff but resident tuition as well and Professor Weaver for his careful reading of the thesis and the time spent in discussing his comments with me. I would also like to thank Professors Thomas P. Carpenter and Michael J. Subkoviak for reading the dissertation.

I am indebted to the many members of the DMP staff and project director James Moser for their assistance in criticizing the items, participating in the validity study and for providing encouragement. I am especially indebted to Marcia Dana, Mary Montgomery, Deborah Stewart (although not a DMP staff member), Gwennyth Trice, and Don Whitaker. The paper would not now be in its present readable form without the assistance of typist and friend, Dorothy Egener.

The support of my sisters, brother, and their families was and continues to be invaluable to me.

To my parents who provided all of their children with a desire to learn and a faith in themselves, I dedicate my thesis.

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS iv

LIST OF TABLES ix

LIST OF FIGURES. xi

ABSTRACT xiii

Chapter

I THE PROBLEM. 1

1.1 Purpose of the Study 1
1.2 Introduction 1
1.3 Background of the Problem. 4
1.4 Questions to be Investigated 5
1.5 Definitions of Terms Used In the Thesis. 6
1.6 Outline of the Thesis. 7

II RELEVANT AND RELATED RESEARCH. 8

2.1 Introduction 8
2.2 Overview of Problem Solving Research 8
2.3 Research in Mathematical Problem Solving 10
2.4 Problem Solving Tasks. 12
2.5 Group Administered Tests of Problem Solving 17
2.6 Summary of Chapter 19

III THE TEST 21

3.1 Chapter Outline. 21
3.2 Introduction 21
3.3 Test Format. 22
3.4 Examples of Superitems 23
3.5 Development of the Test. 28
3.6 Questions to be Investigated 31
3.7 Summary. 35

Chapter page

IV THE RESULTS. 36

4.1 Introduction 36
4.2 The Effect of the Superitem format on Item Response. 36
4.3 Discussion of the Data Related to the Independency of the Items . . 45
4.4 Conditional Probabilities Associated With the Items. 51
4.5 A Discussion of the Conditional Probabilities 53
4.6 Validity 62
4.7 Discussion of the Content Validity . 67
4.8 Reliability. 69
4.9 Discussion of Reliabilities Associated With the Test. 70
4.10 Summary. 74

V SECONDARY ANALYSES OF THE DATA 77

5.1 Introduction 77
5.2 An Item Analysis 77
5.3 The Results of a Cluster Analysis. . 81
5.4 The Results of a Second Clustering Procedure 86
5.5 The Complexity of the Items. 96
5.6 Summary. 98

VI CONCLUSIONS. 99

6.1 Introduction 99
6.2 Summary and Conclusions of the Study 99
6.3 Limitations of the Study 105
6.4 Studies Using the Problem Solving Test. 108
6.5 Significance of the Study. 114
6.6 Implications of Future Research. . . 116

REFERENCES 122

APPENDIX A: PROBLEM SOLVING TEST. 130

APPENDIX B: ITEM DIFFICULTY AND X_{50} 146

APPENDIX C: DIAMETERS OF CLUSTERINGS. 148

APPENDIX D: K-4 TOPICS OF THE DEVELOPING MATHEMATICAL PROCESSES MATHEMATICS PROGRAM 150

LIST OF TABLES

Table		page
3.1	Timetable of Events.	31
4.1	Description of the Tests	37
4.2	Number of Children Taking Each of the Tests. . .	39
4.3	Test Administration Time	40
4.4	The Means, Variances, and Reliability Estimates of the Scales on Each Test Containing the Scale.	42
4.5	t-Tests Values	44
4.6	Conditional Probabilities Associated With the Superitems	52
4.7	Ordering of the Conditional Probabilities of the Superitems	54
4.8	Categorization of the Items on the Basis of Their Conditional Probabilities.	55
4.9	Classification of Items by Judges.	65
4.10	Total Number of Items Placed into Each of the Categories by the Judges	67
5.1	Item Parameters.	82
5.2	Means on C, A, and P Scale for Each Cluster. . .	84
5.3	Conditional Probabilities for Cluster 1.	87
5.4	Conditional Probabilities for Cluster 2.	88
5.5	Conditional Probabilities for Cluster 3.	89
5.6	Conditional Probabilities for Cluster 4.	90
5.7	Categorization of the Items on the Basis of Their Conditional Probabilities by Cluster . . .	91

Table page

5.8 Means on A and P Scale for Each Cluster 93

5.9 Intercorrelations of the Item 96

6.1 Means, Variances, and Reliability Estimates of the Scales for DMP and Non-DMP Children. . . 110

6.2 Means, Variances, and Reliability Estimates of the Scales for Boys and Girls in Grades 6, 7, and 8. 111

6.3 Results of Non-Multiple Choice and Multiple Choice Forms of the Test. 119

LIST OF FIGURES

Figure		page
3.1	Example 1 of Superitems	24
3.2	Example 2 of Superitems	26
3.3	Example 3 of Superitems	27
5.1	Typical Item Characteristic Curve	79
5.2	Range of the Scores Within Each Cluster	85
5.3	Range of the Scores Within Each Cluster	94

DEVELOPMENT OF A TEST OF MATHEMATICAL PROBLEM SOLVING WHICH YIELDS A COMPREHENSION, APPLICATION, AND PROBLEM SOLVING SCORE

DIANA CATHERINE WEARNE

Under the Supervision of Professor Thomas A. Romberg

ABSTRACT

The purpose of the study was to develop a test of mathematical problem solving behavior which provided information about the child's mastery of the prerequisites of the problem solving questions. To provide this additional information, each problem solving question was preceded by two other questions. One question assessed the child's understanding of the information contained in the item stem of the problem solving question and a second question assessed the child's knowledge of an underlying concept of the problem solving question; the first question was referred to as the comprehension question and the second, the application question. The study was concerned with (1) determining if asking multiple questions on the same unit of information affects the response to those questions, and (2) measuring the extent to which the questions preceding the problem solving question were assessing prerequisites of the problem solving question.

To determine if the response to the items was affected by the other items within the superitem, six tests were administered consisting of subsets of the items; three of the tests contained only one type of item and the other three tests contained two types of items. To determine if the differences between the means on the four tests containing each of the scales Comprehension, Application, and Problem Solving were significant, t-tests were computed. The decision to use t-tests rather than an overall test of significance was prompted by a concern for Type II errors. Five of the 18 t-tests computed were significant at the .05 level, two of which also were significant at the .01 level. One of the significant differences involved the Comprehension scale, three the Application scale, and one the Problem Solving scale.

Conditional probabilities were computed for each superitem in an effort to ascertain if the comprehension item was assessing a prerequisite of the application item and if both comprehension and application items were assessing prerequisites of the problem solving item.

Reliability of the test was estimated by two procedures. A Hoyt procedure was used under the assumption the items were independent and a generalized KR-20 in the event the items were interdependent. The Hoyt estimate was .84 and the KR-20 was .79.

The only indicant of validity that could be obtained was the content validity of the test as measured by a panel of judges interested in research in problem solving. The judges were given the items in a random order and asked to classify the items. The judges' classification of the items was compared to the classification of the items on the test and a measure of association was computed between these two classifications; the computed measure of association was .78

Secondary analyses of the data included an item analysis, two cluster analyses, and an examination of the relative complexity of the items.

Chapter I

THE PROBLEM

1.1 Purpose of the Study

The purpose of the study was to develop a test of problem solving behavior for fourth-grade children which also provided information about the child's mastery of the prerequisites of the problems.

1.2 Introduction

There is unanimity among mathematicians and mathematics educators of the importance of developing the problem solving abilities of children. The Cambridge Conference on School Mathematics (1963) indicated that problem material should receive at least one-half of the time and attention of authors of text series. In its recommendations for college preparatory mathematics, the College Entrance Examination Board (1959) urged that throughout the entire sequence of topics to be offered, the mathematics taught be applied to problems, real or of the puzzle type. The National Advisory Committee on Mathematics Education (1975) stressed the importance of problem solving in school mathematics programs.

The K-13 Report Committee of the Ontario Institute for Studies in Education (1971) believed that problem solving should pervade all mathematics curricula. The role of problem solving, as seen by the members of the Report Committee, is to motivate the study of mathematics, to test the understanding of concepts, to stimulate curiosity, to clarify the relationship between mathematics and the world we live in, and to motivate the study and understanding of the structural aspects of mathematics. Obviously, the authors consider problem solving to be the central activity in school mathematics.

Fehr (1967) included the ability to solve problems as one of his three goals for elementary school mathematics. The 37th Yearbook of NCTM (1975) devoted an entire chapter to the topic of problem solving. The authors of the chapter, Nelson and Kirkpatrick, state that problem solving is of basic importance. They assert that the main purpose of instruction in mathematics is to help the child learn how to relate a situation or event that occurs in his life to a mathematical representation of that situation or event and to solve that representation.

In response to all of these recommendations, problem solving has attained some degree of prominence in text series. For example, Ginn's Elementary Mathematics (1972) and Random House's Mathematical Program (1972) both list problem solving as an integral part of their programs.

Another series, Developing Mathematical Processes (DMP) published by Rand McNally in 1975, views problem solving as the vehicle for achieving the goals of its program (Romberg & Harvey, 1969). DMP is a research based individually guided instructional program in elementary mathematics developed by the staff of the Analysis of Mathematical Instruction Project at the Wisconsin Research and Development Center for Cognitive Learning at the University of Wisconsin. The activity approach to learning used by the program is referred to by the authors as learning through problem solving.

There has been some discussion of the type of problems to involve in a mathematics program. A weakness in practical problem solving has been consistently and broadly criticized, most notably by Kline (1973). Kline has been critical of the application problems appearing in mathematics texts as purporting to represent "real life" situations but having little in common with what constitutes real life. Nelson and Kirkpatrick (1975) also have emphasized the importance of "real life" situations. Others believe the "real life" category to be too restrictive and have advocated any problem susceptible for mathematical analysis (Nacome, 1975). However, there is no disagreement on the importance of including problem solving activities in mathematics programs.

Polya's (1962) oft quoted statement concerning the importance of problem solving voices the feelings of virtually all mathematics educators, to wit:

> What is know-how in mathematics? The ability to solve problems - not merely routine problems but problems requiring some degree of independence, judgment, originality, and creativity [Polya, 1962, viii].

In addition to advocating the inclusion of problem solving material in school mathematics program, mathematics educators are actively engaged in research on problem solving behavior, particularly the heuristics of problem solving (see Chapter II). Thus, it would appear that a measure of problem solving behavior is needed in order to determine to what extent the goal of developing problem solving abilities is being attained.

1.3 Background of the Problem

An individually administered test of problem solving behavior has the advantage not only of producing a score but also of providing opportunity to observe the child solving the problems and to ask the child how he solved them or, if he was unable to solve them, to be aware of his path of reasoning and the aspects of the problems which were confusing to him.

However, the constraint of a limited amount of time usually allotted for assessing problem solving behavior necessitates the use of a group administered test. A major difficulty associated with a group administered test is the inability of the examiner to identify the subjects who were unable to solve the problem because they either did not understand the information presented in the question or because they had not mastered the prerequisite concepts of the problem solving question as opposed to those who were unable to solve the problem although possessing the necessary prerequisites.

Studies have reported that the primary factors related to success in problem solving are: reading to note details, understanding of the vocabulary, mastery of the necessary computational skills, and knowledge of the relevant mathematical concepts (Johnson, 1944; Chase, 1964). Treacy (1944) and Alexander (1960) reported that good and poor problem solvers differed in many aspects of reading. Specific instruction in quantitative vocabulary was found by Vanderlene (1964) to be effective in increasing problem solving scores. In another study, Egan and Green (1973) reported that individual differences in specific prerequisite knowledge was more important for "discovery" learning and creative problem solving than for "rule" learning. Norman (1951) found that merely possessing the necessary computational skills did not imply that a child will be able to solve a verbal problem which uses the respective computational skill.

The research reported in this section underscores the possibility than an incorrect response to a problem solving question actually may be the result of the child's lack of familiarity with the vocabulary contained in the question or not possessing the appropriate level of concept attainment necessary for the problem solving question.

1.4 Questions to be Investigated

The proposed test sought to provide information about the child's understanding of the vocabulary and his mastery of the prerequisite concepts. This information was to be obtained by asking a series of three questions rather than a single problem solving question.

The first question assessed the child's understanding of the information contained in the item stem of the problem solving question. The second question assessed the child's knowledge of an underlying concept of the problem solving question. The third question was the problem solving question. The first question was referred to as the comprehension question and the second, the application question. [1](The test is described in detail in Chapter III.)

The format of such a test raises a series of questions; the dissertation describes the manner in which the questions were investigated and the results of the investigations. The questions are as follows:

1. Does asking a series of questions have an effect, either facilitating or debilitating, on the response to the questions; in particular, does the inclusion of a comprehension and an application question have an effect on the response to the problem solving question?
2. How will the reliability be estimated?
3. What type of validity will be obtained?
4. Are correct responses to the comprehension and application questions prerequisite for a correct response to the problem solving question?

1.5 Definitions of Terms Used in the Thesis

The following terms will be used in the thesis:

A *class* is an instructional unit.

An *item* is a question asked on the problem solving test.

A *superitem* is a group of items.

---

[1] An amplification of the definitions may be found on page 63.

A problem situation is a situation which poses a question whose solution is not immediately available, that is, a situation which does not lend itself to an immediate application of some rule or algorithm.

1.6 Outline of the Thesis

Chapter II contains a brief review of the research in problem solving and a description of the categories of tasks and tests used to measure problem solving behavior. The development of the test of problem solving behavior outlined in Section 1.5 is described in Chapter III. Chapter IV contains the results of the initial analysis of the data related to the questions posed in Section 1.5. The secondary analysis of the data is presented in Chapter V and Chapter VI includes a discussion of the implications and limitations of the study.

Chapter II

RELEVANT AND RELATED RESEARCH

2.1 Introduction

This chapter begins with a brief outline of problem solving research, primarily by psychologists, continues with a review of recent research in mathematical problem solving, and concludes with a discussion of some of the instruments used to measure problem solving behaviors. Specifically, Section 2.2 contains an overview of problem solving research. Recent research in mathematical problem solving is described in Section 2.3. A description of some of the problem solving tasks used by the researchers mentioned in Section 2.2 is contained in Section 2.4, and group administered tests of problem solving are reviewed in Section 2.5. Section 2.6 serves to summarize this chapter.

2.2 Overview of Problem Solving Research

Problem solving has received considerable attention from psychologists during the past fifty years. This interest has produced a myriad of seemingly unrelated tasks and an area of research which has been described as one of the most chaotic areas of psychological investigations (Davis, 1966; Coats, Alluisi, & Morgan, 1971).

The variety of research reflects the different theoretical bases of the investigators. The Gestaltists favor the insightful nature of the problem solving process (Duncker, 1945; Wertheimer, 1945; Asher, 1963; Sheerer, 1963). Duncker defined a problem as the situation when a subject has a goal but does not know how this goal is reached. The solution of a problem involves repeated analysis of the current state of the problem and the goal, each step resulting in a transformation of the problem which shortens the gap between the given and the goal.

The information processing theorists view the problem solving process as a search for an appropriate solution from a set of potential solutions (Newell, Shaw, & Simon, 1958; Laughery & Gregg, 1962; Johnson, 1964; Posner, 1965). The means-end analysis, as employed by the information processors, is similar to Duncker's concept of the problem solving process with the exception that the Gestaltists allow the creation of a new method; a means-end heuristic search system finds differences between current and desired situation, finds an operator relevant to each difference, and applies the operator to reduce the difference (Simon & Newell, 1971). The information processors prefer complex problem solving tasks, referred to as task environments, and have concentrated their efforts in chess playing, discovering proofs in logic, and cryptarithmetic (Simon & Newell, 1971).

Other researchers have employed factor analytic techniques in a structure-of-intellect approach (Guilford, 1956; Merrifield, Guilford, Christensen, & Frick, 1962; Dye & Very, 1968). Guilford defines a problem as one in which initial information is given and some goal is more or less specified but the relationship between the given and the goals is not apparent.

The antithesis of the Gestaltists are those who take a learning approach to problem solving (Kendler & Kendler, 1962; Staats & Staats, 1963). This group of researchers would include those concerned with teaching heuristics of problem solving (Covington, 1966; Means & Lorel, 1968; Post, 1968; Lucas, 1972; Jerman, 1973). Other research in this broad category would include those concerned with transfer effects and rigidity (Rees & Israel, 1935; Luchins & Luchins, 1950; Campbell, 1960; Mandler, 1962).

The brief outline of problem solving research contained in this section provides an indication of the breadth of the research that has been and continues to be conducted by psychologists. Some of the investigators were interested in the problem solving process, others the factors associated with success in problem solving, and others who sought to investigate procedures which modify problem solving behaviors. All of the investigators share an interest in this complex human behavior.

2.3 Research in Mathematical Problem Solving

Some of the research in mathematical problem solving was described in Section 1.3. The research presented in that section attempted to identify some of the factors related to success in problem solving in mathematics. The gist of that research is *that familiarity* with the vocabulary used in the problem solving question and possession of the required computational skills are a necessary but not sufficient condition for successful problem solving.

Another group of researchers have investigated the effects of heuristically oriented instruction on problem solving behavior. Bassler, Beers, and Richardson (1975) used two strategies in instructing ninth-grade algebra students to solve verbal problems. One method was derived from Polya and the other was based on Dahmus (1970) and was a direct, pure, piecemeal, and complete translating of the verbal statement into mathematical symbols. Both methods proved useful in teaching the solution of verbal problems although the authors preferred the Polya method. Jerman (1973) also studied the effects of two programs, Covington's Productive Thinking Program (Covington, Crutchfield, & Davis, 1966) and his own Modified Wanted Given Program (Jerman, 1971). He found that teaching problem solving to fifth graders can best be done in a mathematics context using a wanted-given approach. However, if transfer of problem solving skills to other subjects outweighs the importance of higher level of achievement in problem solving in mathematics, then the Productive Thinking Program is better. The authors noted that use of either program was better than nothing; that is, just having the students do problems does not appear to improve their problem solving skills.

Post and Brennan (1976) investigated the effects of various modes of presentation of general heuristic processes on students' problem solving behavior. The results of the study would seem to indicate that formal instruction in a heuristic model suggesting general components of problem solving is not effective in promoting problem solving ability, at least as measured by the tests used in the study with tenth-grade geometry students. Another study which attempted to analyze the effect of heuristic-oriented teaching was reported by Lucas (1972). Lucas investigated the heuristic usage and problem solving performance of first-year calculus students. At two points in time, students were asked to solve a set of mathematical word problems while thinking aloud. The results were analyzed using a modification of a system of behavioral analysis devised by Kilpatrick (1967). Lucas found that heuristicly oriented instruction had a positive effect on changing heuristic strategies and enhancing certain aspects of problem solving performance.

The research reported in this section emphasizes the need for reliable and valid measures of problem solving behavior. Both Jerman (1973) and Post (1976) indicated that there was a lack of available instruments to measure change in problem solving behavior.

2.4 Problem Solving Tasks

The tasks used in the studies reported in Section 2.2 are somewhat more difficult to place into broad categories than the investigators who used them. Ray (1955) described 29 problem solving tasks. The majority of these tasks may be described as of the puzzle or gaming type rather than of the performance experimental measure type. The tasks described by Ray either proved to be so simple that the subject could respond almost immediately or so complex that the investigator was unable to follow the path that led the subject to solve the problem. There was also a tendency of the reseachers to judge the quality of the performance on these tasks, for example, the nonsensical errors of Duncker or the Wertheimer good and stupid solutions.

However, since the time of Ray's (1955) article, there has been a trend toward greater quantification and objectivity in the investigation of problem solving behavior (Coates & Alluisi, 1971). In particular, there has been greater emphasis on being able to obtain information on the entire process of problem solving by designing tasks which yielded this information. A description of some of these tasks is contained in the remainder of this section.

One category of tasks are those consisting of unusual uses of physical objects such as Maier's two-strings, hatrack, or pendulum problems, Katona's matchstick problem or Saustad's ball problem (Maier, 1930; Katona, 1940; Saustad, 1960). In the two-string problem, for example, the subject is presented with two strings dangling from the ceiling too far apart to permit anyone to reach both of them at the same time, or to carry one over to where he can reach the other. There are many other objects in the room including a chair, a stick, additional string, and pliers. The problem is to tie the two pieces of string together.

The fact than an investigator cannot accurately predict from among many alternatives available to a subject which routes he will elect to pursue or what features of the situation will stimulate decisions has led to the criterion of another large group of tasks. The prototype of this task is Lazerte's (1933) envelope test. In this test the subject, at each stage in the problem solving process, selects one of several alternative procedures written on different envelopes. Once he has chosen a procedure, he opens the envelope to find another set of envelopes offering alternatives, and continues in this manner until he reaches a solution. Tab items (Fattu & Mech, 1953; Glaser, Damrin, & Gardner, 1954; Rimaldi, 1955; Buswell, 1956; Rimaldi, Asher, & Burger, 1968) and trouble shooting tasks (Bourne & Battig, 1966) are other examples of this type of task along with medical diagnosis problems (McGuire & Babbott, 1965).

An example of an envelope-type task has been constructed by Neimark and Lewis (1967). They devised a task which they considered to be relatively content free and also provided objective quantitative measures of the quality of the performance, a claim made by all envelope-type tests. The subject was presented an answer sheet containing n numbered patterns each composed of k black or white circles and a problem board in which one of these patterns is concealed under k movable shutters, one over each element. The subject's task is to identify the concealed pattern by uncovering as few of its elements as possible.

An interest in studying rigidity or transfer effects has led to the development of another category of tasks. These tasks consist of a series of problems which are variations of one another and which can be solved by similar techniques followed by problems related in some way to the original series except that either these terminal problems are to be solved by a different technique (for example, the water-jar problems of Luchins and Luchins (1950) or the anagram problems of Reese and Israel (1935) or the terminal problems do not resemble the original series but are solved by similar techniques [for example, Katona's card trick problems (1940)].

There are also a group of tasks which may be called light pattern tasks. These tasks consist of lights which are controlled by switches. The switches may control one, two, or three lights. In the majority of these studies, the subject's task is to produce a specified pattern of lights by selective switching (French, 1954; Duncan, 1963; Davis, 1966). Further examples of these tasks are those involved with the study of the Klein-Four Group (Branca, 1971; Pereira, 1973).

Some of the light pattern tasks contain an additional light to indicate to the subject when he has solved the problem or a phase of the problem (John & Miller, 1958, Gyr, 1960). Still others have a stage model in which the subject must not only determine the sequence of switches which will yield a specified pattern of lights but also the transformation which produces one pattern from another pattern, that is, joins the phases (Alluisi & Coates, 1969; Coates & Alluisi, 1971).

Moore and Anderson (1954) found the calculus of propositions to be a fruitful source of tasks. The characteristics possessed by this source which makes them efficacious for experiments in problem solving are as follows: the mode of symbolication and content may be varied, problems of varying degrees of complexity may be generated, the subject's progress toward solution may be followed, pseudo tasks may be devised, few subjects have had previous experience in the area, and the tasks may be individually or group administered.

Virtually all of the tasks described in this section are individually administered; exceptions to this are the water-jar problems of Luchins and Luchins (1950), some of the envelope tests, and the tasks of Moore and Anderson (19[illegible]). Thus, an investigator interested in assessing problem solving behavior with a limited amount of time available to assess such behavior must consider other instruments.

2.5 Group Administered Tests of Problem Solving

A cursory examination of existing group administered tests which purport to assess problem solving behaviors reveals that the authors of these tests apparently have defined a problem solving situation as a verbal problem, the majority of which are one-step (operation) problems. The operation required to solve the problem is frequently implied by the wording of the problem itself; for example, asking "What is the area of . . .?" or "How much more . . .?"

The Stanford Achievement Test (Harcourt, Brace, & World, Inc., 1964) contains a section entitled "Applications." However, the Directions for Administering, Intermediate I Battery refers to that portion of the tests as measuring problem solving. Examples from this section are as follows:

> 2. Don is delivering papers to earn more money. He had 150 papers to deliver an hour ago. He has delivered 90 of them now. How many are left to deliver?
>
> 25. If two pencils cost 15¢, how many can you buy for 30¢?
>
> 29. Cathy has 4 dolls, Lynn has 5, Sue has 9. What is the average number of dolls they have?

Intermediate I Battery is designed for children in Grade 4 and the first half of Grade 5.

The Modern Math Understanding Test, Form C, Multilevel Edition (SRA, 1966) classifies 12 items as being problem solving or application problems. All of the items may be classified as simple applications or assessing concepts. Examples from this test are as follows:

9. On a certain map a distance of one inch represents 200 miles. If the distance between two towns is 2 1/2 inches on the map, how many miles apart are the towns?

17. Which numeral must be placed in the box to make the following sentence true?

$$6 \times 2 = (\square \times 2) + (1 \times 2)$$

35. The perimeter of this rectangle is _______________. (A rectangle is pictured with 11 inches written on one side and 4 inches on another side.)

Other items in this set refer to the concept of relatively prime numbers, equivalent fractions, and adding the lengths of line segments together.

The Iowa Test of Basic Skills (Houghton, Mifflin Company, 1971) and the Metropolitan Achievement Tests (Harcourt, Brace, & Jovanovich, 1971) also contain problem sets identified as assessing problem solving behavior. The comments made relative to the Stanford Achievement Test and the Modern Math Understanding Test are appropriate for these two tests as well.

An alternative to the standardized tests of problem solving is a test consisting of items which conform to the investigator's definition of problem solving (Kilpatrick, 1967; Zalewski, 1974). Zalewski (1974) explored the feasibility of using a written test to predict seventh-graders' mathematical problem solving achievement as assessed by an interview-coding procedure. The method for validly assessing subject's mathematical problem solving behavior in this study was a thinking aloud procedure. Because of the complexity of the interview and coding schemes, a written instrument which hopefully had high concurrent validty was developed so that it could be used as an alternative to the interview and coding procedures. Although the correlation coefficient between the written and interview test scores did not reach the level established for feasibility (.71), Zalewski hypothesized that a more reliable test might produce the desired level of correlation.

2.6 Summary of Chapter

Section 2.3 contained an overview of the research in problem solving behavior. Some of the recent research in problem solving in mathematics was reviewed in Section 2.4. Section 2.5 contained a description of some of the problem solving tasks used by the researchers described in Section 2.3. A review of the tasks has shown there exist many interesting tasks, interesting for the examiner and for the subject participating in the experiment. However, the majority of these tasks are individually administered and, as such, are not suitable in situations in which a limited amount of time is allocated for assessing problem solving behavior.

Existing standardized tests were reviewed in Section 2.5. These tests failed to meet the criterion established in this study for problem solving tests, that is, the items do not satisfy the definition of problem solving used in this study. (The definition is contained in Chapter I.) Other group administered tests such as those constructed by Zalewski (1974) contain problem solving situations but provide no information on the subject's readiness to solve the problem, that is, no information concerning the subject's mastery of the skills needed to solve the problem.

Chapter III

THE TEST

3.1 Chapter Outline

This chapter describes the development of the problem solving test. Section 3.2 is an introduction to the chapter. The test format is described in Section 3.3. The development of the test is described in Section 3.5 and the procedures for investigating the questions posed in Chapter I are given in Section 3.6. Section 3.7 is a summary of the chapter.

3.2 Introduction

When a child is presented with a problem solving task, one would like to be confident that (1) the child understands the information contained in the task description, and (2) he has the prerequisite skills for the problem solving task. Without this information, one cannot state with any degree of certainty that a child cannot solve a problem because he has responded incorrectly. For example, if the problem solving tasks involves the concept of area and the child is unfamiliar with this concept, then an incorrect response will not necessarily indicate that the child is not capable of solving the problem, merely that he is unfamiliar with the underlying concept.

The test developed for this study attempted to provide information about the child's mastery of the prerequisites for the problem solving question. It was felt that such a test not only would yield more information about the child but also would provide a "truer" measure of his problem solving ability by making it possible to consider only those problems for which the child has demonstrated preparedness.

3.3 Test Format

The test was designed so that it produced three scores: a comprehension score, an application score, and a problem solving score. To accomplish this, the test was composed of groups of items called superitems. Each of the superitems contained a comprehension question, an application question, and a problem solving question. (Examples of superitems are contained in Section 3.4.)

The comprehension question assessed the child's understanding of the information contained either implicity or explicitly in the item stem. The application question assessed the child's mastery of a prerequisite concept or skill of the problem solving question; this question was a fairly straightforward application of some rule or concept to a situation. The third question in the set was the problem solving question.

A problem situation was defined to be a situation which posed a question whose solution was not immediately available, that is, a situation which did not lend itself to an immediate application of some rule or algorithm. An effort was made to construct items which were short and did not appear to be impossible for the child to solve. A guide for the construction might be found in Hilbert's (1906) comment:

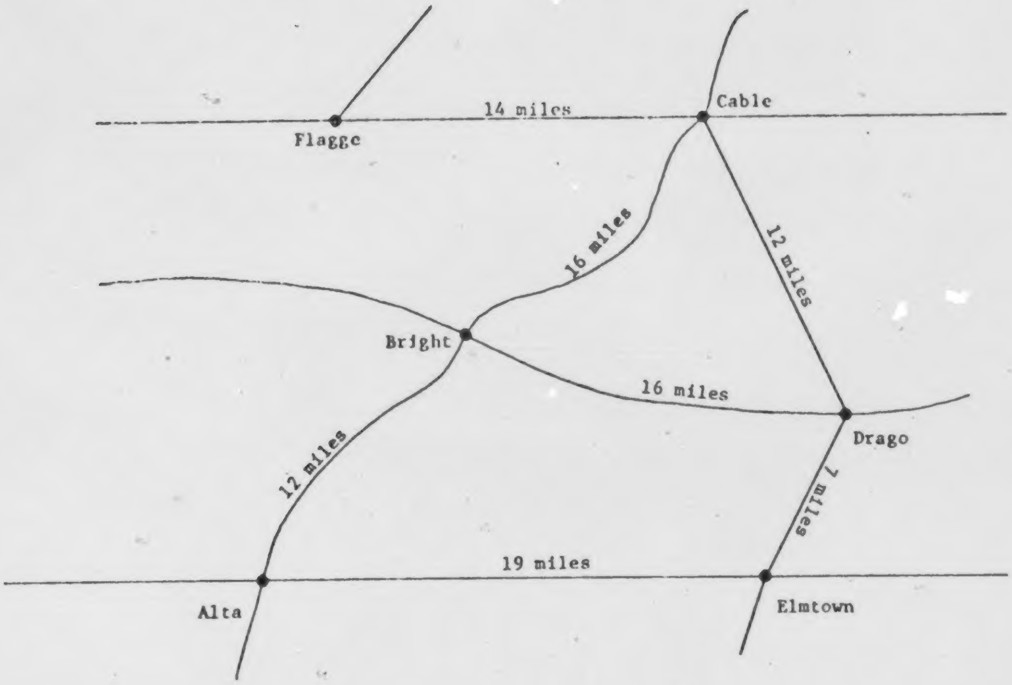
> A mathematical problem should be difficult in order to entice us, yet not completely inaccessible, lest it mock at our efforts. It should be to us a guidepost on the hazy paths to hidden truths and ultimately a reminder of our pleasure in the successful solution [Hilbert, 1906, p. 59].

Although the application and problem solving questions may have referred to a common unit of information (the item stem), the questions were independent to the extent that the response to the application question was not used to respond to the problem solving question. That is not to imply that the underlying concept was not used, merely that the particular response was not used to respond to the problem solving question.

3.4 Examples of Superitems

The first example of a superitem is contained in Figure 3.1. The first question, Question 1-C, is the comprehension question. This question seeks to determine if the child understands the information contained in the map; specifically, to determine if he understands that the numbers contained on the map are distances in miles between towns.

The second question, Question 1-A, is the application question. To respond correctly to this question, the child must be able to


14 miles
Cable
Flagge
16 miles
12 miles
Bright
16 miles
Drago
12 miles
7 miles
19 miles
Alta
Elmtown

The distance from Alta to Bright is:
- 7 miles
- 12 miles
- 16 miles
- 19 miles

The shortest distance from Alta to Drago is:
- through Bright
- through Cable
- through Elmtown
- through Flagge

The sign

BRIGHT	16
ELMTOWN	19

should be placed:
- in Drago
- in Alta
- in Flagge
- in Cable

Figure 3.1. Example 1 of superitems.

read the map, identify the distances referred to by the question, and to correctly add the two distances together.

To respond correctly to Question 1-P, the problem solving question, the child must be able to read the map, add together the two numbers, and find a position on the map which corresponds to the given distance.

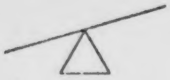
The second example is contained in Figure 3.2. The first question is the comprehension question, Question 2-C. This question is assessing information which is contained implicitly in the problem solving question, that is, how does the balance beam look when it is in balance.

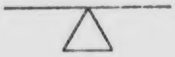
Question 2-A, the application question of this superitem, assesses the child's ability to determine the order relation between two objects on the attribute of weight.

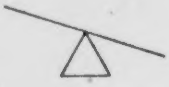
The problem solving question, Question 2-P, assesses other than a routine application of some rule or algorithm. The child must identify that weights A and Z are the lighter weights on the two given balance beams and recognize that the sum of the two lighter weights will be less than the sum of the two heavier weights.

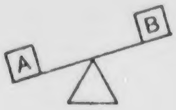
Example 3 is shown in Figure 3.3. Once again, the first question is the comprehension question, Question 3-C. This question asseses the child's understanding of the information presented in the item stem. The application question, Question 3-A, is a direct application of the information contained in the item stem.

Circle the picture that shows the balance beam in balance.









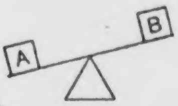
Circle the sentence that tells about the weights A and B .

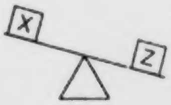
A weighs less than B .

A weighs more than B .

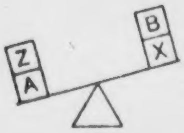
A weighs the same as B .

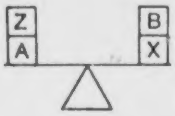
impossible to tell from the picture

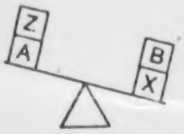




Weights A and Z are put together on one end of the balance beam and B and X are put together on the other end of the balance beam. Circle the picture that shows how the balance beam might look.

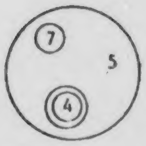


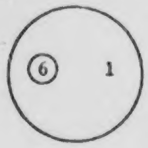




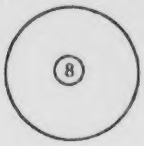
impossible
to tell

Figure 3.2. Example 2 of superitems.

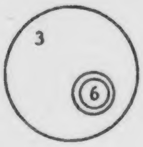
In Circleland, people write 

when they mean 475 and they write 

when they mean 61.

In Circleland, people write 

when they mean 8. TRUE FALSE

What do they mean when they write 

?

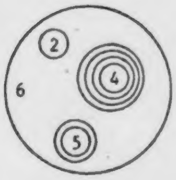
36

63

630

603

306

What do they mean when they write 

?

4,526

40,526

4,562

45,620

45,260

Figure 3.3. Example 3 of superitems.

The problem solving question asks the child to arrive at a generalization of the information given in the item stem; the problem solving question is Question 3-P.

3.5 Development of the Test

A further impetus to the development of a test of problem solving was the desire of the Analysis of Mathematics Instruction (AMI) project staff at the Wisconsin Research and Development Center for Cognitive Learning, developers of the K-6 mathematics program Developing Mathematical Processes (DMP), to include a measure of problem solving as one component of their terminal accountability tests (Romberg, 1974). (Other tests to be included in the terminal accountability tests are measures of concept attainment and computational skills.)

Because the test would ultimately constitute a model for the terminal accountability tests of the DMP program, further constraints were imposed on the application and problem solving questions; to wit: the application question should measure mastery behaviors of the program and the problem solving questions should measure behavior beyond the mastery level of children at the end of the fourth grade. For example, finding the area of a rectangle or a figure composed of rectangles given its dimensions is a mastery behavior for this program by the end of the fourth grade; finding the area of a nonrectangular figure or a figure not composed of rectangles given its dimensions is beyond the mastery level at this point in the program. Thus, finding the area of a rectangle would be an appropriate application question and finding the area of a nonrectangular region would constitute a problem solving question.

A pool of 29 superitems was constructed. These items were reviewed by the AMI staff early in 1975. The reviewers were asked to note if the application questions represented mastery behavior of the program at the stipulated point in time and if the problem solving questions were suitable for the children at this grade level. The reviewers also were encouraged to make qualitative comments about the items and to indicate which items they liked and which they disliked. Some of the reviewers constructed rating scales for the items indicating those which were most suitable and those which were least suitable, others merely made comments about each of the items.

Based upon the reviewers comments, 19 superitems (57 questions) were selected from this pool and were tried out with 8 fourth-grade children. No revisions were deemed necessary after this tryout.

This test of 19 superitems was used in a study by Meyer (1976) in April, 1975. On the basis of the results of this "tryout," three items were deleted. Two of the items were deleted because the children were apparently unfamiliar with the concepts (perimeter and average) at the time of the testing. (This was particularly relevant as the testing time for the present test was in December and, therefore, there was less opportunity for the children to have mastered the concepts.)

A third item was deleted due to poor item parameters for the problem solving question; the beta and biserial correlations values were negative. The underlying concept for this item was writing a fraction to represent a shaded portion of a figure.

A fourth item was altered markedly from the version used by Meyer (1976) to the present form. The original item had asked the child (in the application question) to determine the number of diagonals of a quadrilateral; the comprehension question had asked the child to identify a line as being or not being a diagonal of a figure. The problem solving question gave a table with the number of diagonals for figures with 3-7, 9, and 10 sides and asked the child to determine the number of diagonals for an eight-sided figure. It was felt that the underlying concept for this problem solving question was not diagonals but completing a number pattern. Thus, the new item's comprehension and application questions were number patterns and the problem solving question consisted of the same number pattern as before but with no physical referrent.

Another pool of items was constructed and then reviewed by the same staff of people. The reviewers were again asked to be candid in their remarks and to indicate which items they felt were appropriate for the test. Six items were selected from this new pool, based on the reviewers comments, and the new test of

22 superitems was tried out with 7 fourth-grade children. The only result of this tryout was a wording change on one question.

A timetable of the events described in the preceding paragraphs is given in Table 3.1.

TABLE 3.1

TIMETABLE OF EVENTS

Activity	Date
Review by AMI staff of first set of items	February, 1975
Eight person tryout of first version of the test	March, 1975
Tryout of first version of the test	April, 1975
Review by AMI staff of second set of items	September, 1975
Six person tryout of second version of the test	October, 1975
Administration of test	December, 1975

3.6 Questions to be Investigated

A test composed of superitems yields more information than one containing only the problem solving questions; however, structuring the test in this manner raises some questions and areas of concern. These questions were posed in Chapter I. The manner in which the questions were investigated is described in this section. An amplification of the procedures for investigating the questions and the results of these investigations is contained in Chapter IV.

The questions are as follows:

1. Does asking a series of questions have a facilitating or debilitating effect on the response to the questions; in particular, does the inclusion of a comprehension and an application question have an effect on the response to the problem solving question?

It is possible that asking the child a comprehension question may focus his attention on one aspect of the problem and asking the application question may suggest an approach to solving the problem. Focusing the child's attention in this manner may either facilitate or debilitate the problem solving process. Burch (1953), in a study with elementary school children, found that asking the child questions concerning what was known and what was to be found frequently merely fragmented the problem. Thus, it was necessary to determine the effect the three questions (comprehension, application, and problem solving) contained in each superitem had on one another.

To measure the effect, six tests consisting of subsets of items from the total test were to be constructed. One test would contain only the comprehension items, another the application items, and a third, the problem solving items. Three other tests would also be constructed, each containing two types of items: one test would contain the comprehension and the application items, another the comprehension and the problem solving items, and the third would contain the application and the problem solving items. To measure the effect of the questions upon one another, mean scores for each of the three categories of items, comprehension, application, and problem solving, would be compared across the four instruments containing the item category. For example, the mean application score from the test containing only application items, the test containing comprehension and application items, the test containing application and problem solving items, and the test containing all three types of items would be compared to see if there was any significant differences among the mean scores. Thus, it would be possible to determine if the items have an effect upon each other.

2. How will the reliability of the test be estimated?

If the results of the procedures outlined in Question 1 indicate the questions do not have an effect upon one another, then the Hoyt procedure for estimating the reliability of a test would be appropriate. However, if the results indicate the responses to the items are not independent, then the Hoyt procedure would produce an inflated estimate of the reliability of the test.

In the event of this occurrence, the result of the process described in Question 1 indicating the response to the questions are interdependent, then the unit for estimating the reliability will not be the individual items but rather the superitems and a generalized KR-20 will be used to estimate the reliability. This latter procedure was suggested by Cureton (1965),

3. What type of validity will be obtained?

The difficulty with the criterion of validity is derived from the lack of a universally accepted definition of problem solving behavior and, therefore, no established criteria against which the tasks may be validated. The only indicant of validity which can be obtained is the content validity of the tasks as judged by those interested in research on problem solving.

To determine the content validity, a panel of judges would be given the definitions of the three categories of items (comprehension, application, and problem solving) and the items in a random order and would be asked to classify each of the items. The judges' classification of the items then would be compared to the classification of the items on the test and a measure of association would be computed between the judges' and the classification of the items on the test.

4. To what extent is the model for the test supported by the test results?

It had been hypothesized that a correct response to the comprehension and application items was a prerequisite to a correct response to the problem solving item and that a correct response to the comprehension item was a prerequisite to a correct response to the application item. To measure the extent to which the items actually fit the model, conditional probabilities would be computed. The conditional probabilities which were to be computed are as follows:

Prob(comprehension item correct | application item correct)

Prob(comprehension item correct | problem solving item correct)

Prob(application item correct | problem solving item correct)

Prob(comprehension and application items correct | problem solving item correct)

3.7 Summary

Section 3.3 described the test format. Examples of superitems were described in Section 3.4. The development of the test was described in Section 3.5 and the procedures for investigating the questions posed in Chapter I were outlined in Section 3.6.

Chapter IV

THE RESULTS

4.1 Introduction

This chapter contains the results of the initial analysis of the data. Further analyses of the data, including the results of a cluster analysis, are contained in Chapter V.

Section 4.2 contains the data related to the question of the independency of the items on the test with a discussion of the analysis in Section 4.3. The conditional probabilities associated with the superitems are reported in Section 4.4 and examined in Section 4.5. The procedures followed in obtaining the content validity of the test are outlined in Section 4.6 and discussed in Section 4.7. The reliability of the test and the scales associated with the test are reported in Section 4.8 and a discussion of the reliabilities is contained in Section 4.9. The chapter is summarized in Section 4.10.

4.2 The Effect of the Superitem Format on Item Response

Of fundamental interest was the effect of the comprehension, application, and problem solving questions upon each other; that is, the effect of asking multiple questions on the same unit of information upon the response to those questions. The inclusion of the comprehension and application questions provides more information than a test containing only the problem solving questions. However, if including the additional questions has an effect upon the response to the problem solving questions, then the strength of the problem solving test is diminished.

To determine the effect of the items upon one another, six tests consisting of a subset of the original set of items were constructed. Three of the tests contained only one of the three types of items; the remaining three tests contained two of the three types of items. Thus, there were seven tests, the six tests containing a subset of the items and the complete test. The content of the tests is described in Table 4.1.

TABLE 4.1

DESCRIPTION OF THE TESTS

	Type of Items		
	Comprehension Items	Application Items	Problem Solving Items
C	x		
A		x	
P			x
CA	x	x	
CP	x		x
AP		x	x
CAP	x	x	x

Each of the tests containing a subset of the items was administered to approximately 50 children. To minimize teacher effect as much as possible, each test was administered in four classes. Each of the four classes was in a different school with the exception of two classes in the same school who took the C test. The children in each class were randomly assigned to one of two instruments. The complete test was administered either by the investigator or by a graduate student who had been instructed in the manner in which to respond to the children's questions; the other six tests were administered by the investigator. The number of children taking each of the seven tests is given in Table 4.2.

The population for this study consisted of fourth-grade children living in Wisconsin who had been using the Developing Mathematical Processes (DMP) mathematics program for at least one and one-half years. The children who took the complete test (CAP) either attended one of four schools in a city of population 40,000 which is a suburb of a large city or they attended one of two schools located in a suburb of a medium sized city. The children taking the subtests attended schools in one of the following locations: a city of population 50,000, a small town, a suburb of a medium sized city, and one school in that city.

It was somewhat difficult to arrive at an administration time for the six tests containing a subset of the items. The administration time of the complete test is 45 minutes. However,

TABLE 4.2

NUMBER OF CHILDREN TAKING EACH OF THE TESTS

	Test							
School	C	A	P	CA	CP	AP	CAP	Total
1	10	14			13	11		48
2		12	12	14		11		49
3		8	7	11	9			35
4	19	13	25	10	14	12		93
5	9	3	5	7	10	12		46
6							90	90
7							53	53
8							68	68
9							37	37
10							31	31
11							38	38
Total	38	50	49	42	46	46	317	588

it is not possible to merely apportion time based upon the number of items. Two assumptions affected the administration time. Those assumptions are listed below.

1. It was assumed the problem solving items require more response time than do the application items which, in

turn require more response time than do the comprehension items.

2. Since the items frequently shared the same item stem, it was assumed that the child did not have to process two pieces of information and, hence, would not need sum of the times needed to respond to the items individually to respond to the items in pairs.

Using these assumptions, the administration times chosen were as follows: 20 minutes for the C test, 25 minutes for the A test, 30 minutes for the P test, and 35 minutes for the CA, CP, and AP tests. The administration times are summarized in Table 4.3.

TABLE 4.3

TEST ADMINISTRATION TIME

Test	Administration Time
C	20 minutes
A	25 minutes
P	30 minutes
CA	35 minutes
CP	35 minutes
AP	35 minutes
CAP	45 minutes

As noted previously, the test contains three types of items; for the remainder of this paper, these categories of items will be referred to as scales.

The number of subjects, means, variances, and reliability estimates for the Comprehension, Application, and Problem Solving scales on each of the tests containing them are reported in Table 4.4.

To determine the significant differences between the mean scores of the scales on the four tests containing each scale, t-tests were computed. Normally one is not justified in computing multiple t-tests as t-tests carried out on all pairs of means must necessarily extract redundant information from the data and, as a result, a complicated pattern of dependency emerges between the tests. In fact, even if the tests were independent of one another, it would still be true that the probability of at least one significant result by chance alone would be $1 - (1 - .05)^n$ where .05 is the level of significance and n is the number of t-tests. For this study, the probability of one significant difference by chance alone would be $1 - (1 - .05)^{18}$ (approximately .60) among the t-tests related to the Comprehension, Application, or Problem Solving scales or $1 - (1 - .05)^{21}$ (approximately .66) if all the tests are taken into consideration. Thus, it appears likely that one significant result will occur by chance. The corresponding probability for nonindependent comparisons is difficult to ascertain (Hays, 1973, p. 593).

TABLE 4.4

THE MEANS, VARIANCES, AND RELIABILITY ESTIMATES OF THE SCALES ON EACH TEST CONTAINING THE SCALE

Scale	Test	Number	Mean	Variance	Reliability
C	C	38	15.95	6.97	.49
	CA	42	17.02	7.54	.58
	CP	46	15.93	10.06	.65
	CAP	317	15.53	9.74	.63
A	A	50	11.08	17.22	.78
	CA	42	10.50	15.43	.76
	AP	46	8.61	15.09	.75
	CAP	317	10.01	13.15	.71
P	P	49	4.24	11.02	.74
	CP	46	2.93	4.37	.49
	AP	46	3.46	7.81	.69
	CAP	317	3.30	6.13	.60
CA	CA	42	27.52	38.74	.82
	CAP	317	25.54	37.41	.80
CP	CP	46	18.87	18.96	.69
	CAP	317	18.82	22.53	.73
AP	AP	46	12.07	38.86	.84
	CAP	317	13.31	31.26	.80
CAP	CAP	317	28.84	62.18	.84

In addition, Cochran and Cox (1957, p. 74) have pointed out that for the extreme case in which the largest and smallest of three or more means are compared, a large difference due entirely to chance variation may be observed. The probability of spurious differences occurring increases with the number of means compared. For example, if the number of means is four (the number of means associated with any one of the Comprehension, Application, or Problem Solving scales), the observed value of t for the largest difference will exceed the tabled five percent value approximately 20 percent of the time.

In this study, the intent of comparing the means was to determine if the items had an effect on one another. Consequently, it was more serious to neglect to identify a significant difference than to identify more significant differences than actually exist. Or, stated another way, a Type II error was of greater consequence than a Type I error. The results of the t-tests are summarized in Table 4.5.

Five of the 21 t-tests were significant at the .05 level, two of which were also significant at the .01 level. There was one significant difference among the comprehension scores, three among the application scores, and one among the problem solving scores.

The significant difference among the comprehension scores was between the mean comprehension score on the CA test and the mean comprehension score on the CAP test. This difference was significant at the .01 level. All three of the significant differences among the application scores involved the application score on the AP test.

TABLE 4.5

t-TESTS VALUES

C scores

	CA	CP	CAP
C	-1.78	.03	.91
CA		-1.73	-3.25 **
CP			.80

A Scores

	CA	AP	CAP
A	.69	3.01 **	1.72
CA		3.27 *	.77
AP			-2.30 *

P Scores

	CP	AP	CAP
P	2.32 *	1.24	1.90
CP		01.03	-1.09
AP			.37

CA Scores		CP Scores		AP Scores	
	CAP		CAP		CAP
CA	1.02	CP	.07	AP	-1.32

* p < .05

** p < .01

The mean application socre on the AP test was significantly different from the mean application score on the A, CA, and CAP test. The difference between the mean application score on the A and AP test was significant at the .01 level.

The one significant difference among the problem solving scores was between the mean problem solving score on the P test and the mean problem solving score on the CP test. This difference was significant at the .01 level.

4.3 Discussion of the Data Related to the Independency of the Items

The data related to the independency of the items was presented in the previous section. Five significant differences were found among the 21 t-tests computed. One of the significant differences involved the Comprehension scale, three the Application scale, and one the Problem Solving scale.

Two hypotheses may be advanced to account for these differences: (1) the items have an effect, under certain conditions, upon one another, and (2) the children did not have the same amount of time to respond to a particular group of items on all of the tests containing that group of items.

It may be hypothesized that the higher mean comprehension score on the CA test was due to a facilitating effect of the application items. This effect could occur if the child responded to the comprehension item after responding to the application item, a retroactive effect, or if the inclusion of the application items had a stimulating effect upon the response to the other items on the test, the comprehension items. The CAP test also contained the application items but it could be asserted that the facilitating effect of the application items on this test was counterbalanced by a debilitating effect of the problem solving items or that a retroactive effect does not take place in the presence of the problem solving items. However, if a debilitating effect was produced by the problem solving items, one would think this effect would also have appeared on the CP test which was not the case. A possible argument still remains that the debilitating effect of the problem solving items upon the comprehension items only takes place in the presence of the application items; although such an intricate dependency is possible, it does not seem likely.

Another possible explanation for the significant difference lies in the administration times of the tests. It is possible that the children taking the CA test had more time to respond to the comprehension items than the children taking the other three tests containing the comprehension items. The CA and CP tests both had an administration time of 35 minutes; however, the application items are assumed to be less difficult than the problem solving items, and thus, the children may well have had more time to respond to the comprehension items on the CA test than on the CP test. Although the mean comprehension score on the CA test was not significantly different from the mean comprehension score on the CP test, the mean on the CP test differed from the mean comprehension score on the CAP test by .40 points. The administration time for the CAP test was 45 minutes, ten minutes longer than the administration time of the CA test but the CAP test included the problem solving items; the problem solving items are assumed to be the most difficult and, hence, require more response time than do the other two types of items.

Therefore, of the two possible explanations for the significant differences proposed in the preceding paragraphs, the most reasonable appears to be associated with the administration times of the tests.

A similar pattern of mean scores occurred among the application scores except that in this case, the mean score of one test was lower than the mean scores of the other three tests. The mean application scores on the A, CA, and CAP test ranged from 10.01 to 11.08; the mean application score on the AP test was 8.61. The mean application score on the AP test was significantly lower than the mean application score on all the other tests containing the application items.

One possible explanation for this difference is that the problem solving items on the AP test had a debilitating effect upon the response to the application items. However, this same effect did not occur on the CAP test although it could be argued that the effect of the problem solving items on the CAP test was tempered by the effect of the comprehension items on the application items. This interdependency may be in effect, but it does not appear likely.

Once again, another explanation for the difference lies in the administration time allotted for the tests. The CA and AP tests both had an administration time of 35 minutes. However, the AP test contains the two most difficult types of items and, consequently, the children would have had less opportunity to respond to the application items on this test (AP) than they did on the CA test. A similar argument could be put forth for the significant difference between the mean application score on the A test and the AP test as the children had 25 minutes to respond to the application items on the A test and only 35 minutes to respond to both the application and the problem solving items on the AP test. The CAP test contained only the comprehension items in addition to the items contained on the AP test. Although the comprehension items require the least amount of responding time, the administration time of the CAP test was ten minutes longer than that of the AP test thereby providing more time for the children to respond to the application and problem solving items than was provided on the AP test.

Thus, of the two hypotheses advanced to explain the significant differences existing between the mean application score on the AP test and the mean application socres on the other three tests containing the application items, the most reasonable appears to be the effect of the time allotted to respond to the items on the test.

There was only one significant difference involving the problem solving mean score. The significant difference was between the mean problem solving score on the P test and on the CP test. As before, the two possible hypotheses put forth to explain the difference are that the comprehension items had a debilitating effect on the problem solving items or that the administration times of the tests account for the differences. For this significant difference, it is possible that both may have been a contributing factor.

The investigator noted during the administration of the CP test that some children were disturbed by the disparity in difficulty of the two types of questions. Although the children had been cautioned that some of the items on the test were a great deal more difficult than others, the children apparently forgot these instructions and were disturbed over being able to answer some items with relative ease while others remained obscure to them. This may have affected their willingness to pursue the problem solving items for any length of time.

Another possible explanation for the significant difference may lie in the administration time of the tests. The children had 30 minutes on the P test to respond to the problem solving items and 35 minutes to respond to the problem solving and the comprehension items on the CP test. Therefore, the children taking the P test may have had more opportunities to attempt problem solving items than did the children on the CP test. However, the children taking the AP test had the same amount of testing time as the children taking the CP test and the difference between the problem solving score on the AP test was not significantly different from the mean problem solving score on the P test. This result indicates that a reason for the significant difference may be the effect of the comprehension items on the problem solving items when just the two categories of items appear on the same instrument and that this is a more likely reason than the administration times of the tests.

As noted in Section 4.2, the probability of arriving at a spuriously significant difference if 21 independent t tests are made is .66. In this study the comparisons are clearly not independent and thus increases the probability that one of the significant differences discussed in this section is spuriously significant. This argument is given added weight by noting that when four means are compared, the difference between the largest and the smallest of the four means may be spuriously significant at the .05 level twenty percent of the time (Cochran & Cox, 1957).

This concludes the investigation of the independency of the items. The data related to determining to what extent the first two items in each superitem were assessing prerequisites of the problem solving item is reported in the following section.

4.4. Conditional Probabilities Associated With the Items

The model for the superitem on the test is the initial two items are assessing prerequisite behaviors of the problem solving item. The data related to determining to what extent the superitems fit the model is presented in this section.

It had been hypothesized that a correct response to the comprehension item was a prerequisite to responding correctly to the application item and that a correct response to the problem solving item. To determine to what extent this was true, the following conditional probabilities were computed:

Prob(Comprehension item correct | Application item correct)

Prob(Comprehension item correct | Problem Solving item correct)

Prob(Application item correct | Problem Solving item correct)

Prob(Comprehension and Application | Problem Solving item correct)
items correct

The values for the Prob(Comprehension item correct | Application item correct) ranged from .63 to 1.00 with a mean conditional probability of .86. The Prob(Comprehension item correct | Problem Solving item correct) varied from .52 to 1.00 with a mean probability of .82. The mean conditional probability of Prob(Application item correct | Problem Solving item correct) was .67; the values of the probabilities ranged from 0 to .95. The Prob(Comprehension and Application items correct | Problem Solving items correct) ranged from 0 to .95 with a mean probability of .62. The conditional probabilities for the items are listed in Table 4.6.

TABLE 4.6

CONDITIONAL PROBABILITIES ASSOCIATED WITH THE SUPERITEMS

Item	$P(c \mid a)^1$	$P(c \mid p)^2$	$P(a \mid p)^3$	$P(c \cap a \mid p)^4$
1	.98	.99	.83	.83
2	.98	.99	.92	.91
3	.90	.85	.85	.74
4	.78	.80	.66	.63
5	.95	.97	.94	.92
6	.81	.74	.83	.66
7	.79	.77	.90	.71
8	.96	1.00	.80	.80
9	.91	.91	.83	.78
10	.70	.76	.47	.41
11	.78	.71	.76	.62
12	.85	.77	.62	.52
13	.63	.77	0	0
14	.93	.97	.80	.77
15	1.00	.90	.29	.29
16	.72	.59	.24	.24
17	1.00	1.00	.95	.95
18	.91	.97	.83	.79
19	.72	.52	.13	.11
20	.94	.86	.64	.64
21	.75	.74	.67	.50
22	.98	.94	.86	.86

[1] Conditional probabilites: Prob(comprehension item correct | application item correct)

[2] Prob(comprehension item correct | problem solving item correct

[3] Prob(application item correct | problem solving item correct)

[4] Prob(comprehension and application items correct | problem solving item correct)

An ordering of the superitems based upon their conditional probabilities for each of the four conditional probabilities of interest is contained in Table 4.7.

For a superitem to fit the model, all four of the conditional probabilities should reflect the agreement. Superitems were divided into three categories based upon their conditional probabilities; the divisions were arbitrarily selected by the investigator. If .75 is considered to be a minimum conditional probability at which to consider a superitem to be acceptable, then ten of the superitems satisfied this criterion, that is, ten of the superitems reported a conditional probability of at least .75 for all four of the conditional probabilities computed. There were seven additional superitems all of whose conditional probabilities were at least .50; these superitems were considered to be marginally acceptable. Five superitems did not have all four conditional probabilities of at least .50. Table 4.8 lists the specific superitems numbers of the items fitting into each of the categories Acceptable, Marginally Acceptable, and Unacceptable.

4.5 A Discussion of the Conditional Probabilities

There were five superitems whose conditional probabilities placed them in the Unacceptable category, superitems 10, 13, 15, 16, and 19. Four of these superitems contained problem solving items which ranked among the six most difficult problem solving items on the test. The

TABLE 4.7

ORDERING OF THE CONDITIONAL PROBABILITIES OF THE SUPERITEMS

Conditional Probability	Superitems Numbers			
	$P(c \mid a)$	$P(c \mid p)$	$P(a \mid p)$	$P(c \cap a \mid p)$
.90 - 1.00	1,2,3,5,8, 9,14,15,17, 18,20,22	1,2,5,8,9 14,15,17, 18,22	2,5,7,	2,5,17
.80-.89	6,12	3,4,20	1,3,6,8,9, 14,18,22	1,8,22
.70-. 79	4,7,10,11, 16,19,21	6,7,10,11 12,13,21	11	3,7,9,14,18
.60 - .69	13		4,12,20,21	4,6,11,20
.50 - .59		16,19		12,21
.40 - .49			10	10
30 - .39				
.20 - .29			15,16	15.16
.10- .19			19	19
.00 - .09			13	13

TABLE 4.8

CATEGORIZATION OF THE ITEMS ON THE BASIS
OF THEIR CONDITIONAL PROBABILITIES

Category	Conditional Probability	Superitems with All Four of the Conditional Probabilities at That Level
Acceptable	.75 - 1.00	1, 2, 3, 5, 8, 9, 14, 17, 18, 22
Marginally Acceptable	.50 - .74	4, 6, 7, 11, 12, 20, 21
Unacceptable	.00 - .49	10, 13, 15, 16, 19

difficulty levels of the problem solving items for superitems 10, 13, 15, 16, and 19 were .05, 04, .07, .05, and .26, respectively; the mean difficulty level of all the problem solving items was .15. These five item superitems also contained application items which were among the six most difficult on the test. The difficulty levels of the application items of these superitems were .43, .05, .08, .09, and .16 for superitems 10, 13, 15, 16, and 19 respectively; the mean difficulty level of all the application items on the test was .46. Therefore, for four of the five superitems, at most seven percent of the children responded correctly to the problem solving items and for four of the five items, at most 16 percent of the children responded correctly to the application item. These low difficulty values may have contributed to the low conditional probabilities in that the probabilities were based on very small samples. For example, only 13 children responded correctly to the problem solving item of superitem 13 out of a population of 317 children.

An examination of each of the five superitems in the Unacceptable category provides further clues to the low conditional probabilities reported. A discussion of each of the five items is contained in the following paragraphs.

Superitem 10

The four conditional probabilities associated with this superitem are as follows:

$$P(c \mid a) = .70$$

$$P(c \mid p) = .76$$

$$P(c \mid p) = .47$$

$$P(c \cap \mid a \mid p) = .41$$

The application item for this superitem was as follows:

> A parking lot has room for 8 rows of cars with 9 cars parked in each of these rows. How many cars can be parked in the parking lot?

Forty-three percent of the children responded correctly to this item. The behavior required to respond correctly was recognizing that the solution could be found by correctly computing eight groups of nine.

The problem solving item was as follows:

> In another parking lot, trucks are parked. Each truck takes the space of 3 cars. There are 12 trucks in the parking lot and it is completely full. If there were 4 rows in the parking lot, how many cars could be parked in each row?

The problem solving item involved two steps, grouping and partitioning (in either order). For one solution one would first find the number of cars in the parking lot (12 groups of 3) and then the number of cars in each row ($36 \div 4$). To solve another way one would first partition ($12 \div 4$) and then multiply that result by three to obtain the solution.

Only five percent of the children responded correctly to the problem solving item; 70 (22%) of the children responded with 36 indicating they either had not read the entire question and/or were misled by the application item which asked the children to determine the number of cars in the parking lot. In either event, the superitem is a poor one, and thus, for reasons other than its low conditional probabilities, the superitem belongs in the Unacceptable category.

Superitem 13

The four conditional probabilities associated with this item are as follows:

$$P(c \mid a) = .63$$

$$P(c \mid p) = .77$$

$$P(a \mid p) = 0$$

$$P(c \cap a \mid p) = 0$$

This superitem was concerned with the concept of area. The application item asked the child to compute the area of a rectangle (pictured) given its dimensions; five percent of the children responded correctly to this question. (The dimensions of the rectangle were four feet by six feet.) The problem solving item asked the child to determine the area of a rectangular floor not covered by a rug. A drawing was provided to illustrate the problem. Four percent of the children responded correctly to this item, but none of the 13 children who responded correctly to the problem solving item responded correctly to the application item. This would be difficult to understand were it not that since the item was multiple choice, the child could select an answer without having to make any computations. It seems unlikely that a child who could compute the area of a rectangle and subtract that area from another given area to arrive at an answer would be unable to compute the area of another rectangle particularly if both of these areas involved multiplying a one-digit number by a one-digit number. The most likely reason for this occurrence would appear to be that the children guessed correctly. Therefore, it is hypothesized that the item does fit the model but due to the reasons outlined above, the conditional probabilities failed to show this.

Superitem 15

The conditional probabilities associated with this item are as follows:

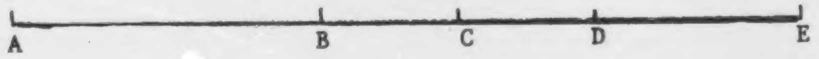
$$P(c \mid a) = 1.00$$

$$P(c \mid p) = .90$$

$$P(a \mid p) = .29$$

$$P(c \cap a \mid p) = .29$$

The item was as follows:



The distance from A to E is 23 cm.

The distance from D to A is 17 cm.

The distance from B to E is 14 cm.

C is half-way between B and D.

What is the distance from A to D? (Comprehension item)

What is the distance from A to B? (Application item)

What is the distance from B to C? (Problem Solving item)

The application item was assumed to be a prerequisite behavior for the problem solving item. This does not imply one must find the distance AB when finding the distance from B to C but that the behavior of subtracting one of the given distance from another was the prerequisite tested. One might respond to the problem solving item by first computing the distance DE and then subtract this distance from BE to find the length of BD and then, finally, BC. Therefore, one step in the solution is to subtract one given distance from another. Rather than ask for one of these distances, the distance AB was requested.

Only eight percent of the children responded correctly to the application item and seven percent to the problem solving item. The conditional probabilities of correctly responding to the application item given that one has responded correctly to the problem solving item was .29. The item was a not a multiple-choice item and so it would appear that there was no relationship, or at most a weak relationship, between the application and problem solving items and thus the item does belong in the Unacceptable category.

Superitem 16

The conditional probabilities associated with this item were:

$$P(c \mid a) = .72$$

$$P(c \mid p) = .59$$

$$P(a \mid p) = .24$$

$$P(c \cap a \mid p) = .24$$

The problem solving item for this superitem was as follows:

> Another ship has 70 passengers. If all of the passengers and crewmen are sitting in the lifeboats, there are 7 passengers and 2 crewmen in each lifeboat. How many crewmen are on the ship?

It was thought that the prerequisite behavior, other than conceptualizing the problem, was determining the number of groups and then finding a given number of those groups, that is, if there are seven passengers in each boat and 70 passengers altogether, then there are ten boats and if two crewmen are in each boat, then there would be 10 groups of 2 or 20 crewmen.

The application item referred to a ship as well. The item stem was as follows:

> The dining room of a ship has 14 tables. One-half of the tables have 6 chairs at each table and the other half of the tables have 4 chairs at each table.

The application item was:

> What is the largest number of people that can sit in the dining room at one time?

To respond to this question, the child had to know that one-half of 14 is seven and be able to find seven groups of six and seven groups of four and add them together. Thus, the behavior consisted of three steps: finding one-half of fourteen, determining how many are in a given number of groups of a given number, and adding two numbers together.

Five percent of the children responded correctly to the problem solving item and nine percent to the application item. The probability of a child responding correctly to the application item given that he responded correctly to the problem solving item was .24. It is possible that this superitem does not fit the model; however, too few of the children responded correctly to the superitem parts to be certain of this conclusion.

Superitem 19

The conditional probabilities associated with this superitem were as follows:

$$P(c \mid a) = .72$$

$$P(c \mid p) = .52$$

$$P(a \mid p) = .13$$

$$P(c \cap a \mid p) = .11$$

Twenty-four percent of the children responded correctly to the problem solving item of this superitem but only 16 percent responded correctly to the application item. Consequently, it is not surprising that the probability of a correct response to the application item given that the problem solving item was responded to correctly was .13. It would appear this item was correctly placed in the Unacceptable category.

Five superitems were placed in the Unacceptable category because of their associated conditional probabilities. The conclusion of this section is that two of these five superitems, superitems 13 and 16, may have been placed in this category for reasons other than the superitems not fitting the model.

4.6 Validity

Due to a lack of an established criteria against which the tasks could be validated, the only indicant of validity that could be obtained was the content validity of the tasks as judged by a panel of experts.

A test may be said to have content validity if it measures something which a group of authorities asserts it does measure. The American Psychological Association and the American Educational Research Association in their joint publication Standards for Educational and Psychological Test and Manuals defines content validity as follows:

> Content validity is demonstrated by showing how well the content of the test samples the class of situations or subject matter about which conclusions are to be drawn [APA, 1966, p. 12].

Thus, evaluating the content validity of a test is tantamount to evaluating the adequacy of a definition.

Six judges were selected on the basis of their familiarity with the DMP materials and an interest in the research on problem solving. A constraint on the development of the superitems for the test had been that the items be suitable for fourth-grade children in the DMP program. (This was discussed in Chapter III.) Thus, the judges had to be familiar with the content of the DMP program in order to be able to judge whether or not a behavior required represented the routine application of a concept or algorithm (application item) or if the behavior required represents other than a routine application of a concept or algorithm (problem solving item).

The judges were given the definitions of the three categories of items (comprehension, application, and problem solving) and the items from the test in a random order and asked to classify the items as representing a comprehension item, an application item, or a problem solving item. The definitions given the judges are as follows:

> Comprehension Item: This item assesses the child's understanding of the information contained either implicitly or explictly in the item stem. When the item is assessing information contained implicitly in the item stem, it may be thought of as assessing the understanding of the definition of a basic concept underlying the situation.

Application Item: This item is a fairly straightforward application of some rule or concept to a situation. This item may be thought of as assessing what is considered to be mastery behavior in DMP at the end of the fourth grade (Topic 65).

Problem Solving Item: A problem situation is defined to be a situation which poses a question whose solution is not immediately available, i.e., a situation which does not lend itself to an immediate application of some rule or algorithm. This item may be thought of as assessing behavior beyond the mastery level of DMP at the end of the fourth grade.

The judges' classification of the items was compared to the classification of the items on the test. The mean proportion of judges agreeing with the test classification was .84. The mean proportion of agreement with the test classification of comprehension items was .89, the mean agreement on the application items was .78, and the mean proportion of judges agreeing with the problem solving classification was .84. Table 4.9 lists the proportion of judges agreeing with the test classification for each item on the test.

There were nine items, out of a total of 66, to which fewer than two-thirds of the judges agreed with the test classification of the items; half of the judges agreed with the test classification on five of these nine items. One of the nine items was a comprehension item, five were application items, and the remaining three were problem solving items. The tendency, when disagreeing with the application category, was to rate the item as problem solving. The comprehension items were rated as application items and the problem solving as application items when disagreeing with the classification of the items on the test for these nine items.

A measure of association was computed between the judges' classification of the items and the classfication of the items

TABLE 4.9

CLASSIFICATION OF ITEMS BY JUDGES

Super-item	Test Classification	Judges' Classification C	A	P	Proportion of Judges' Agreeing with Test Classification
1	C	6	0	0	1.00
	A	5	1	0	.17
	P	0	5	1	.17
2	C	5	1	0	.83
	A	0	5	1	.83
	P	0	4	2	.33
3	C	5	0	1	.83
	A	2	4	0	.67
	P	0	0	6	1.00
4	C	6	0	0	1.00
	A	1	3	2	.50
	P	0	0	6	1.00
5	C	5	1	0	.83
	A	0	4	2	.67
	P	0	0	6	1.00
6	C	5	1	0	.83
	A	0	6	0	1.00
	P	0	2	4	.67
7	C	6	0	0	1.00
	A	0	4	2	.67
	P	0	0	6	1.00
8	C	5	0	1	.83
	A	0	6	0	1.00
	P	0	1	5	.83
9	C	6	0	0	1.00
	A	0	6	0	1.00
	P	0	0	6	1.00
10	C	6	0	0	1.00
	A	0	6	0	1.00
	P	0	0	6	1.00
11	C	6	0	0	1.00
	A	0	6	0	1.00
	P	0	0	6	1.00

TABLE 4.9 (Continued)

Super-item	Test Classification	Judges' Classification C	A	P	Proportion of Judges' Agreeing with Test Classification
12	C	6	0	0	1.00
	A	0	6	0	1.00
	P	0	0	6	1.00
13	C	6	0	0	1.00
	A	0	6	0	1.00
	P	0	2	4	.67
14	C	5	1	0	.83
	A	0	6	0	1.00
	P	0	2	4	.67
15	C	6	0	0	1.00
	A	0	3	3	.50
	P	0	0	6	1.00
16	C	4	2	0	.67
	A	0	2	4	.33
	P	0	0	6	1.00
17	C	4	2	0	.67
	A	0	3	3	.50
	P	0	0	6	1.00
18	C	6	0	0	1.00
	A	0	6	0	1.00
	P	0	0	6	1.00
19	C	6	0	0	1.00
	A	1	4	1	.67
	P	0	2	4	.67
20	C	6	0	0	1.00
	A	1	5	0	.83
	P	0	0	6	1.00
21	C	5	1	0	.83
	A	0	6	0	1.00
	P	0	0	6	1.00
22	C	3	3	0	.50
	A	1	5	0	.83
	P	0	3	3	.50

on the test. The strength of the association was computed to be .77; the index used was Cramer's Statistic ϕ' (Hays, 1973, p. 745). The number of items placed into each of the categories by the judges is shown in Table 4.10.

TABLE 4.10

TOTAL NUMBER OF ITEMS PLACED INTO EACH OF THE CATEGORIES BY THE JUDGES

Test Classification	Judges' Classification C	A	P	Total
C	118	12	2	132
A	11	103	18	132
P	0	21	11	132

4.7 Discussion of the Content Validity

Six judges classified the items. One judge's classfication differed from the test classification on six of the 66 items. Three other judges differed on eight of the items. One judge differed 16 times and the remaining judge differed from the classification on the test on 18 occasions. Thus, four judges differed a total of 30 times and the remaining two judges differed a total of 34 times. The four judges who differed the fewest number of times had all functioned as writers of the DMP program for a minimum of two years. Of the two judges who differed the greatest number of times, one had been a writer for only one year and had been actively involved with the revision of the K-2 materials rather than in writing the 3-4 materials; the test contains more material related to the 3-4 component than the K-2 component of the program. The sixth judge, the judge who differed 18 times with the test's classifications, had been associated with the DMP program since its inception but not as a writer of curriculum materials. It is quite possible that the familiarity gained through writing the materials may have made the judges more aware of what constituted mastery behavior of the program than the two judges who were not involved with writing the 3-4 materials. It is to be remembered that the classification of the application and problem solving questions rests upon mastery objectives of the program.

No one judge consistently rated items as being in a higher or lower category than the classification of the items on the test. However, the two comprehension items rated as representing a problem solving situation were done so by the same judge.

Two judges made no "errors" in categorizing the comprehension items, that is, they did not differ from the classification of the items on the test, and another judge made no "errors" in rating the problem solving items.

In conclusion, the mean proportion of judges agreeing with the test classification of the comprehension items was .89, the mean agreement on the application items was .78, and the mean proportion of judges agreeing with the problem solving classification was .84. The computed measure of association between the judges' classification of the items and the classification of the items on the test was .77.

4.8 Reliability

Two estimates of reliability of the test were computed. The Hoyt reliability and the KR-20 reliability. The Hoyt reliability was computed under the assumption that the items on the test were independent. However, if an indication of dependency was found, then the Hoyt estimate would be an inflated estimate of the reliability of the test. To satisfy this possible contingency, a generalized KR-20 was also computed for the test. This procedure was suggested by Cureton (1965) when discussing the problems associated with computing reliabilities of a test consisting of superitems.

$$r = 1 - \frac{MS_{\text{persons x items}}}{MS_{\text{persons}}} \qquad \text{(Hoyt, 1941)}$$

$$r = \frac{n}{n-1}\left[1 - \frac{\sum^{n} \sigma_i^2}{\sigma^2_{\text{test}}}\right] \qquad \text{(Cureton, 1965)}$$

where n is the number of superitems, σ_i^2 is the score-variance on each of the superitems and σ^2_{test} is the variance of the scores on all the superitems.

Hoyt reliability estamates were also computed for each of the six tests consisting of a subset of the items as well as for each the scales, Comprehension, Application, and Problem Solving, contained in the six tests. Thus, a reliability estimate of the comprehension scores was obtained on the C, CA, CP, and the CAP tests. The reliabilities for the Comprehension scale ranged from .49 on the C test to .65 on the CP test. The reliabilities for the Application scale ranged from .71 on the CAP test to .78 on the A test. The reliability estimates for the Problem Solving scale ranged from .49 on the CP test to .74 on the P test.

The reliabilities computed for the CA, CP, and AP tests were .82, .69, and .84, respectively. The reliabilities of these same scales on the complete test (CAP) were .80, .73, and .80, respectively. The reliabilities are given in Table 4.4.

A discussion of these reliabilities is contained in the Section 4.9.

4.9 Discussion of Reliabilities Associated With the Test

The question of independence of the items was discussed in Section 4.3. The conclusion reached in this section was that the items did not have any effect upon one another and the effects that were noted were artifacts of the administration time or, in the case of the one significant difference on the Problem Solving scale, of placing two particular categories of items together on one test. Thus, it may be assumed that the reliability estimate of .84 as reported by the Hoyt does not represent an inflated estimate of the reliability of the test.

The difference between the more conservative estimate of .79 as obtained by using a generalized KR-20 and the Hoyt estimate of .84 may be attributed to the fact that when the items are placed into groups of three, as they were to form the superitems, variance among the items which had been considered as part of the item-covariance in the Hoyt estimate now constitute part of the item variance. Consequently, the total variance of the test is reapportioned with a greater proportion of the variance now being part of the item variance rather than the covariance of the items. This results in the ratio of item variance to total variance being larger and, hence, a lowering of the reliability estimate.

Reliabilities were also computed for each scale, Comprehension, Application, and Problem Solving on each test containing the scale; estimates of reliability were also obtained for the CA, CP, and AP tests.

The four reliabilities associated with the Comprehension scale on the tests containing the scale varied from .49 on the C test to .65 on the CP test. The associated variances ranged from 6.97 on the C test to 10.06 on the CP test. The low reliability estimates are to be expected after noting the variances.

The reliability estimates associated with the Application scale ranged from .71 on the CAP test to .78 on the A test; the variances from 13.15 on the CAP test to 17.22 on the A test.

The range of reliability estimates reported for the Problem Solving scale was from .49 on the CP test to .74 on the P test;

the associated variances ranged from 4.37 on the CP test to 11.02 on the P test. The reliability estimates are a reflection of the skewness of the distribution of scores on this scale. The range of the problem solving scores on the CAP tests was from 0 to 15 with a mean of 3.30 and a standard deviation of 2.48. The mean proportion of children responding correctly to the 22 problem solving items on the CAP test was .15.

The reliability estimates for the CA, CP, and AP tests were .82, .69, and .84, respectively. The reliabilities of these scales on the CAP test were .80, .73, and .80 for the CA, CP, and AP tests, respectively.

One way of increasing the reliability estimates of the Comprehension and Problem Solving scales would be to increase the variability in the responses to the items. However, in both cases to do so would violate the definition of these categories. The purpose of the comprehension item was to assess the child's understanding of the information contained either implicitly or explicitly in the item stem. Thus, one would expect a relatively high degree of success on this item. The range of difficulty levels associated with the comprehension item was from .35 to .98; the mean item difficulty was .71. An example of a comprehension item to which virtually all of the children responded correctly is the comprehension item of superitem 2; 96% of the children responded correctly to this item. The item (see Figure 3.2) asked the child to determine the number of miles between two towns, a distance which was given on the map. The purpose of the comprehension items is to determine if the child understands that the numbers on the map represent distances in miles between towns. Therefore, this item is an appropriate comprehension item. To assess a more complex behavior would transcend the definition of the item type.

Traditionally, investigators of problem solving behavior have not been particularly concerned with the reliability of their tasks. It is apparent that the investigators consider the task itself to be of more importance than any associated index of consistency. However, when administering a test to a group of children, one would like to be reasonably confident that the results are reliable, that is, that if the children took the test again, the same ordering of children would occur.

A problem situation was defined to be a situation which poses a question whose solution is not immediately available, that is, a situation which does not lend itself to an immediate application of some rule or algorithm. This definition permits a wide latitude of situations whose only constraint is that the behavior required to solve the problem is not mastery behavior of the DMP program at the end of the fourth grade and yet at the same time be a problem for which the children have the prerequisite behaviors, conceptual or computational.

Reliability of the Problem Solving scale would increase if the items were less difficult. However, raising the reliability at the expense of deleting challenging questions would not be desirable. Lord makes the following comment about reliability which is particularly apt:

> . . . maximizing reliability may sometimes be an undersirable goal. For example, a subset of factual items in an achievement test may yield a more reliable score than the total set of items. This can happen, for example, if the other items involve such hard-to-measure but important traits as reasoning ability and creative thinking [Lord & Novick, 1974, p. 344].

4.10 Summary

Section 4.2 was concerned with the independency of the items on the test. It was of interest to determine if asking multiple questions on the same unit of information affects the response to the questions. To ascertain this, the three item types were administered alone, in pairs, and all three together. Consequently, there were four mean scores which could be computed for the three types of items, comprehension, application, and problem solving. Five significant differences were found; one significant difference was among the comprehension scores, three among the application scores, and one among the problem solving scores.

The conclusions reached in Section 4.3, when discussing the results of Section 4.2, was that the items were independent and the significant differences that were noted either were the result of the administration times of the tests or of the placement of two particular categories of items together on the same instrument.

The conditional probabilities computed were reported in Section 4.4 and examined in Section 4.5. Four conditional probabilities were computed for each superitem. A superitem was deemed to be acceptable if all four of the conditional probabilities were at least .75; 10 superitems were in this category with an additional six considered to be marginally acceptable. Five superitems were considered to be unacceptable in this regard; unacceptable indicated at least one of the conditional probabilities were less than .50. These five superitems were examined with the conclusions that three of the five belonged in the unacceptable category and that not enough information, due to the small number of children responding correctly to the items associated with the superitems, was available about the other two superitems to make a judgment.

Section 4.6 contained a discussion of the content validity of the test. A measure of association was obtained between the categorization of the items on the test as perceived by a panel of judges and the classification of the items on the test. This association was found to be .77.

Section 4.7 contained a discussion of the procedure for estimating the content validity of the test. The conclusion reached was that the judges tended to agree with the classification of the items (comprehension, application, and problem solving) on the test and the two judges who disagreed most frequently with the classification on the test were the two judges least familiar with the material contained in the 3-4 component of the program. This is not to say that these two judges were unfamiliar with this (3-4) section of the program, merely that the other judges may have had a greater familiarity with the 3-4 component, a familiarity gained through actual writing of the activities.

The reliabilities associated with the test were given in Section 4.8 with a discussion of them in Section 4.9. Two reliability estimates were computed for the test. A Hoyt was computed under the assumption that the items on the test were independent and a generalized KR-20 in the event that the items were not independent. The Hoyt estimate was .84 and the KR-20 was .79. The unit for computing the Hoyt estimate was the item and the superitem was the unit for the KR-20 estimate. Due to the conclusions of Section 4.3, that is, that the items are independent, the .79 estimate was considered to be a conservative estimate and the Hoyt estimate was felt to better represent the reliability of the test.

CHAPTER V

SECONDARY ANALYSES OF THE DATA

5.1 Introduction

This chapter contains the results of the second phase of the data analysis. The initial analyses, described in Chapter IV, were concerned with the investigation of the research questions posed in Chapter I. The second phase of the data analysis includes an item analysis, a cluster analysis, and an examination of the relative complexity of the three types of items on the test.

5.2 An Item Analysis

No treatise on a test is complete without a discussion of an item analysis for that test. The analysis for this test was performed using the Generalized Item and Test Analysis Program (GITAP) (Baker, 1963). This procedure yields the following parameters: item difficulty, biserial correlation, X_{50}, and β.

The item difficulty is the proportion of children responding correctly to the item. The biserial correlation (item-criterion correlation) is obtained by hypothesizing the existence of a continuous latent variable underlying the dichotomy imposed in scoring the item. It is assumed that the distribution of measures in the sample for which binary values are given is actually normal but that at some point in the distribution a separation has been made with those cases lying above the point being assigned a score of 1 (correct) and those below a score of 0 (incorrect). One may think of the biserial correlation as a sample measure of association for the item score and the total test score, or, under the assumption the scores are normally distributed and the regression of total test score on the item score is linear, as an estimate of the product moment correlation between the item score and the total test score.

The X_{50} and β are also obtained by hypothesizing the existence of a continuous latent variable underlying the dichotomy imposed in scoring the item. The regression of the hypothetical item score on the criterion is called the item characteristic curve. The X_{50} is the point on the criterion scale, given in standard deviation units, corresponding to the median of the item characteristic curve and β is the reciprocal of the standard deviation of the item characteristic curve. One may think of β as the slope of the item characteristic curve at the point X_{50} although this is technically not true. For example, if the normal ogive is fitted to the item response data, the slope of the curve at the point X_{50}, the point of inflection on the normal ogive item characteristic curve, is $\beta/\sqrt{2\pi}$. Figure 5.1 is an illustration of a typical item characteristic curve. The biserial correlation and β are related in the following manner, under the assumption that the criterion is normally distributed:

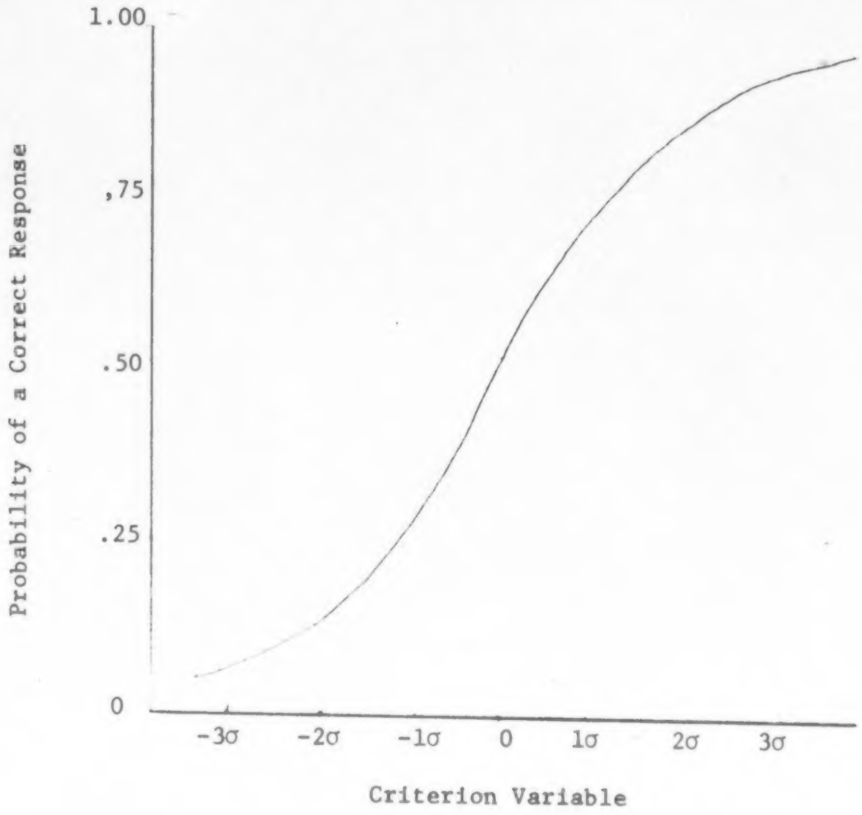


Figure 5.1. Typical item characteristic curve.

$$\beta = \frac{r_{bis}}{\sqrt{1 + r_{bis}^2}}$$

[Lord, 1974, p. 378]

The generally accepted criteria for a "good" item is β and biserial correlation values of at least .30 (Harris, 1968). Obviously, the higher the value of β, the greater the slope of the item characteristic curve which indicates the item is discriminating more clearly. Thus, an item to which everyone responded correctly would be judged a "poor" item. Although one would like high biserial correlations and a high value for β, it is theoretically possible for them to be too high. If, for example, an item had a biserial correlation of 1.00, it would be superfluous to have another such item with the same difficulty level as a person would respond the same to both items. Similarly, an increase in the value of β, while the values of other parameters remain fixed, could lower the precision of an estimation. This is described as the attenuation paradox (Loevinger, 1954). Therefore, an ideal test should have values of varying sizes for β.

Eight items of the 66 on the test had values for β which were less than the desired level of .30. Five of these eight items were computation items, one was an application item, and two were problem solving items. Of these eight items, the value of β for four of the items was between .25 and .29; these values indicate the items are marginally acceptable. All four of these items were computation items. The β value for the comprehension item of superitem 1 (Figure 3.2) was .94; this is not an unusual value for an item to which 98% of the children responded correctly. The value for β of the application item of superitem 13, the area problem discussed in Section 4.5, was -.04. Superitem 13 was the superitem in which none of the children who responded correctly to the problem solving item responded correctly to the application item. Therefore, it is not surprising that the application item would correlate poorly with the total test score. The two low values for β reported for the problem solving items were .20 and .21 for superitems 19 and 20, respectively.

The items with low values for β had biserial correlations values of approximately the same size. The additional item with a low biserial correlation was the comprehension item of superitem 9; the biserial correlation reported for this item was .28 indicating the item is marginally acceptable. The β and biserial correlation values are given in Table 5.1.

In conclusion it may be said that all but four of the items on the test had values for β which were acceptable (at least .30) or marginally acceptable (.25 - .29).

5.3 The Results of a Cluster Analysis

A clustering procedure was used to discover if there was any structure (natural arrangements of children into homogeneous groups)

TABLE 5.1

ITEM PARAMETERS

	β			r_{bis}		
Item	C	A	P	C	A	P
1	.04*	.35	.45	.04*	.33	.41
2	.34	.54	.71	.32	.48	.58
3	.34	.68	.75	.32	.56	.60
4	.28*	.44	1.34	.27*	.40	.80
5	.32	.50	.66	.30	.45	.55
6	.25*	.47	.63	.24*	.43	.53
7	.49	.88	1.17	.44	.66	.76
8	.30	.50	1.71	.28*	.45	.86
9	.40	.54	1.71	.37	.48	.86
10	.47	.79	.37	.43	.62	.35
11	.29*	.51	.65	.28*	.45	.54
12	.26*	.51	.44	.26*	.45	.40
13	.43	-.04*	.39	.39	-.04*	.36
14	.81	.77	.82	.65	.61	.63
15	.59	.60	.60	.51	.51	.51
16	.63	.60	1.21	.51	.53	.77
17	1.29	.96	.63	.79	.69	.53
18	1.18	.92	.81	.76	.68	.63
19	.98	.38	.20*	.70	.35	.19*
20	1.12	.56	.21*	.75	.49	.21*
21	.94	.81	.35	.69	.63	.33
22	.87	.41	.75	.66	.38	.60

*poor questions

inherent in the data. Investigators using the test may correlate measures of interest with any of the three scores produced by the test. However, additional insight may be gained by examining the correlations of specific groups of children with identifiable characteristics in terms of the test.

A Wards clustering procedure (Johnson, 1967) was used. The Wards procedure is a maximum method clustering, that is, the value of the clustering is the maximum diameter of the clusters produced. At any step in the clustering process, the distance from an object in Cluster j to an object in Cluster k is the diameter of the cluster which is the union of Clusters j and k. This procedure appeared to produce four clusters. The diameters of the 15 largest clusterings is included in Appendix C.

Cluster 1 consisted of 91 children whose mean comprehension, application, and problem solving scores were all below the grand means; this group was designated as the (-) group. The mean scores of the 112 children in the second cluster were approximately at the mean for all three scores; this group was called the (0) group. Cluster 3 consisted of 99 children whose mean scores were above the grand means for all three categories of items; this group was called the (+) group. The remaining cluster in this clustering, Cluster 4, consisted of 15 children whose application and problem solving scores were two standard deviations above the mean for the entire group of children; it was impossible to be two standard deviations above the mean for the comprehension scores. Cluster 4 was designated as the (++) group. The mean scores for each of the clusters is given in Table 5.2 and the range of the scores within a cluster is shown in Figure 5.2.

TABLE 5.2

MEANS ON C, A, AND P SCALE FOR EACH CLUSTER

Cluster	Number of Subjects	Means C	Means A	Means P
1 (-)	91	11.98	6.30	1.59
2 (0)	112	15.96	9.30	2.35
3 (+)	99	17.72	13.17	4.91
4 (++)	15	19.33	17.00	10.01
Total Population	317	15.53	10.01	3.30

Conditional probabilities for the items were examined for the children in each of the four clusters. The number of items fitting the model was fewer in the (-) group than in the (++) group; only five superitems were acceptable in the (-) group, indicating that all four conditional probabilities were at least .75, and 16 were acceptable in the (++) group. Ten items were acceptable or marginally acceptable for Cluster 1, 14 for Cluster 2, 19 for

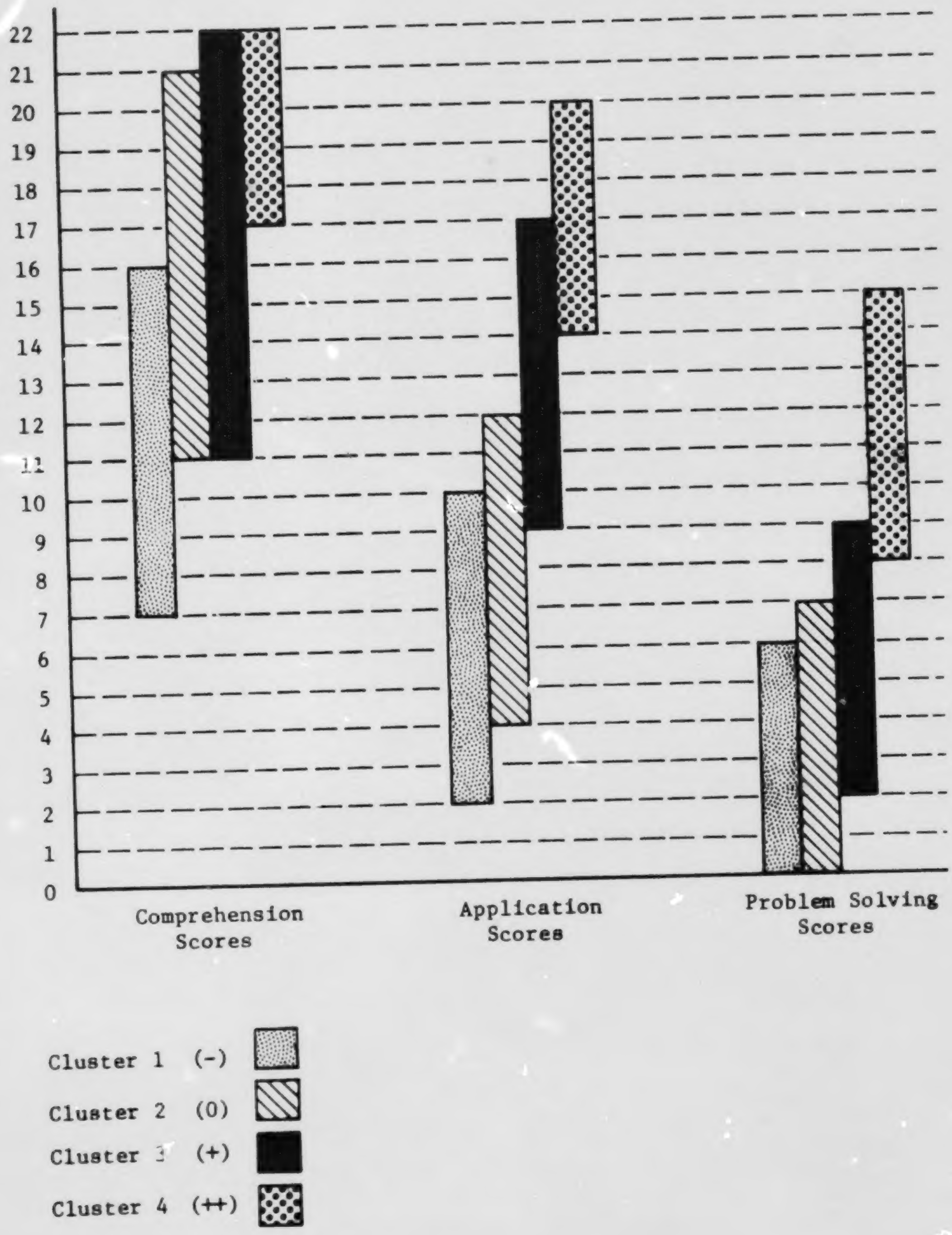


Figure 5.2 Range of the scores within each cluster.

Cluster 3, and 21 for Cluster 4. The conditional probabilities for the superitems in each cluster are contained in Table 5.3 through Table 5.6; the number of superitems fitting the categories of Acceptable, Marginally Acceptable, and Unacceptable for each cluster are contained in Table 5.7.

A superficial explanation for this trend towards more of the superitems fitting the model as the scores of the children increase may be that the conditional probabilities reflect correct responses to more of the items, that is, if the children are responding correctly to virtually all of the items, then the conditional probabilities would necessarily be close to one. However, the mean problem solving score for Cluster 4 (++) was 10.01 out of a possible score of 22; this would indicate that even the children in the highest category were not responding correctly to the majority of the problem solving items.

Another explanation for the trend towards more of the items fitting the model as the score of the children increase is that the children are not responding as erratically to the items which were multiple choice format in the (+) and (++) groups as in the (-) and the (0) group.

5.4 The Results of a Second Clustering Procedure

A second clustering procedure was performed, this time without regard to the comprehension scores. Two factors prompted the decision to examine the data without the comprehension scores: the

TABLE 5.3

CONDITIONAL PROBABILITIES FOR CLUSTER 1

	Item Numbers			
Conditional Probability	P(c | a)	P(c | a)	P(a | p)	P(c ∩ a | p)
.90 - 1.00	1,2,3,5,8,15, 17,18,22	2,5,10,22	1,2,5,8,13, 17,18,22	5,22
.80-.89	12,13,14	3	3,9	2
.70 - .79	6,9,20	1,7	11	1,3
.60 - .69	11	9,17	12,14,20	17
.50 - .59	4,10	6,11,12,18	4,6,7	6,9,11,18
.40 - .49	7		19	12
.30 - .39	21	14,20,21	21	20
.20 - .29	19	4		4,7
.10 - .19		8		
.00 - .09	16	13,15,16,19	10,15,16	8,10,13,14, 15,16,19,21

TABLE 5.4

CONDITIONAL PROBABILITIES FOR CLUSTER 2

Conditional Probability	Item Numbers			
	P(c | a)	P(c | p)	P(a | p)	P(c ∩ a | p)
.90 - 1.00	1,2,3,5,8,9, 15,17,20,22	8,17	1,2,5,8,9, 13,14,15,16, 17,18,20	8,17
.80 - .89	6,13,14,18	1,7,9	3,4	1,5
.70 - .79	4,7,12,19,21	2,3,14	6,7,12,21, 22	2,3,9,14
.60 - .69	10	4,6,11,18 22	10,11	4,22
.50 - .59	11,16	21		6,7, 11,18
.40 - .49			19	
.30 - .39		12		21
.20 - .29		15		12,15
.10 - .10		19		19
.00 - .09		10,13,16,20		10,13,16.20

TABLE 5.5

CONDITIONAL PROBABILITIES FOR CLUSTER 3

Conditional Probability	Item Numbers			
	P(c | a)	P(c | p)	P(a | p)	P(c ∩ a | p)
.90 - 1.00	1,2,5,8,9,11 14,15,17,18, 20,22	2,5,6,17,20	1,2,3,5,8,9, 10,14,15,17, 18	2,5,17
.80 - .89	3,4,7,12,16 19	1,3,7,9,10 14,18,20,22	4,12,20,22	1,3,9,10,14 18,20,22
.70 - .79	6,10,21	11,12,21	6,7,11,21	6,7
.60 - .69		4,8		4,8,11,12
.50 - .59			13,19	21
.40 - .49				
.30 - .39			16	
.20 - .29	13	15		15
.10 - .19		19		
.00 - .09		13,16		13,16,19

TABLE 5.6

CONDITIONAL PROBABILITIES FOR CLUSTER 4

Conditional Probability	Item Numbers P(c | a)	P(c | p)	P(a | p)	P(c ∩ a | p)
.90 - 1.00	1,2,4,5,6,7,8, 9,10,12,13,14, 15,16,17,18, 20,21,22	1,2,3,4,5,6, 7,9,10,11,12, 17,18,21,22	2,4,5,7,8, 9,10,12,13, 14,15,17,18, 19,21,22	2,5,7,9,10,12, 17,18,21,22
.80 - .89	11	8,14	1,6,11	1,4,6,8,11,14
.70 - .79	19		16	
.60 - .69	3	15.20	20	15,20
.50 - .59		16,19	3	3, 16, 19
.40 - .49				
.30 - .39				
.20 - .29				
.10 - .19				
.00 - .09		13		13

TABLE 5.7

CATEGORIZATION OF THE ITEMS ON THE BASIS
OF THEIR CONDITIONAL PROBABILITIES BY CLUSTER

		Number of Items with all Four Condtional Probabilities at that Level			
Category	Probability	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Acceptable	.75 - 1.00	5	5	12	16
Marginally Acceptable	.50 - .74	5	9	6	5
Unacceptable	.00 - .49	12	8	4	1

conditional probabilities and the length of the test. The mean conditional probabilities of responding correctly to a comprehension item given that the application response was correct is .86; if the four lowest conditional probabilities are omitted, the mean conditional probability is .90. Consequently, one may be reasonably certain that a child would respond correctly to the comprehension item if he responded correctly to the application item of that superitem.

The second factor influencing the decision to examine the data for structure without the comprehension scores was the length of the test. The test consists of 66 items to be administered within a 45 minute time period. Deleting the comprehension items would provide additional response time for the application and problem solving items without increasing the administration time. Deleting the comprehension items is a fairly serious step as the items serve to indicate if the child has understood the information contained in the item stem, and incidentally, to provide the child with items which are less difficult at regular intervals. Due to these two factors, it was of interest to see if a clustering procedure would produce meaningful groups if the comprehension items were omitted and if so, to what extent the groups differed from those produced by the initial clustering process.

The initial clustering process produced four groups, three of similar size, 91, 99, and 112, and one small group of 15. The second clustering procedure also appeared to yield four clusters; however, one of these groups was almost one-half of the total group, 147 out of 317 children. Two other groups were approximately the same size, 82 and 72, and one small group of 16.

The four clusters produced may be designated in a manner similar to those formed in the initial clustering process. Cluster A consisted of children whose mean application and problem solving scores were below the grand mean; this group will be designated as the (-*) group. Cluster B consisted of children whose mean application and problem solving scores were similar to the mean scores for the entire group of children; this group will be called the (0*) group. The children in Cluster C had mean application and problem solving scores which were above the mean; this group will be called the (++*) group. The mean scores of the children in the clusters is contained in Table 5.8 and the range of the scores within a cluster is shown in Figure 5.3.

TABLE 5.8

MEANS ON A AND P SCALE FOR EACH CLUSTER

Cluster	Number of Subjects	Means A	Means P
1 (-*)	147	7.10	1.52
2 (0*)	82	10.71	3.28
3 (+*)	72	13.61	5.44
4 (++*)	16	17.00	10.00
Total Population	317	10.01	3.30

The change in the groups from the initial clustering to the second clustering was as follows:

1. There were 147 children in Cluster A, 87 of these were in Cluster 1 and 60 in Cluster 2.
2. Cluster B consisted of 82 children; four of these were in Cluster 1, 50 in Cluster 2 and 27 in Cluster 3.

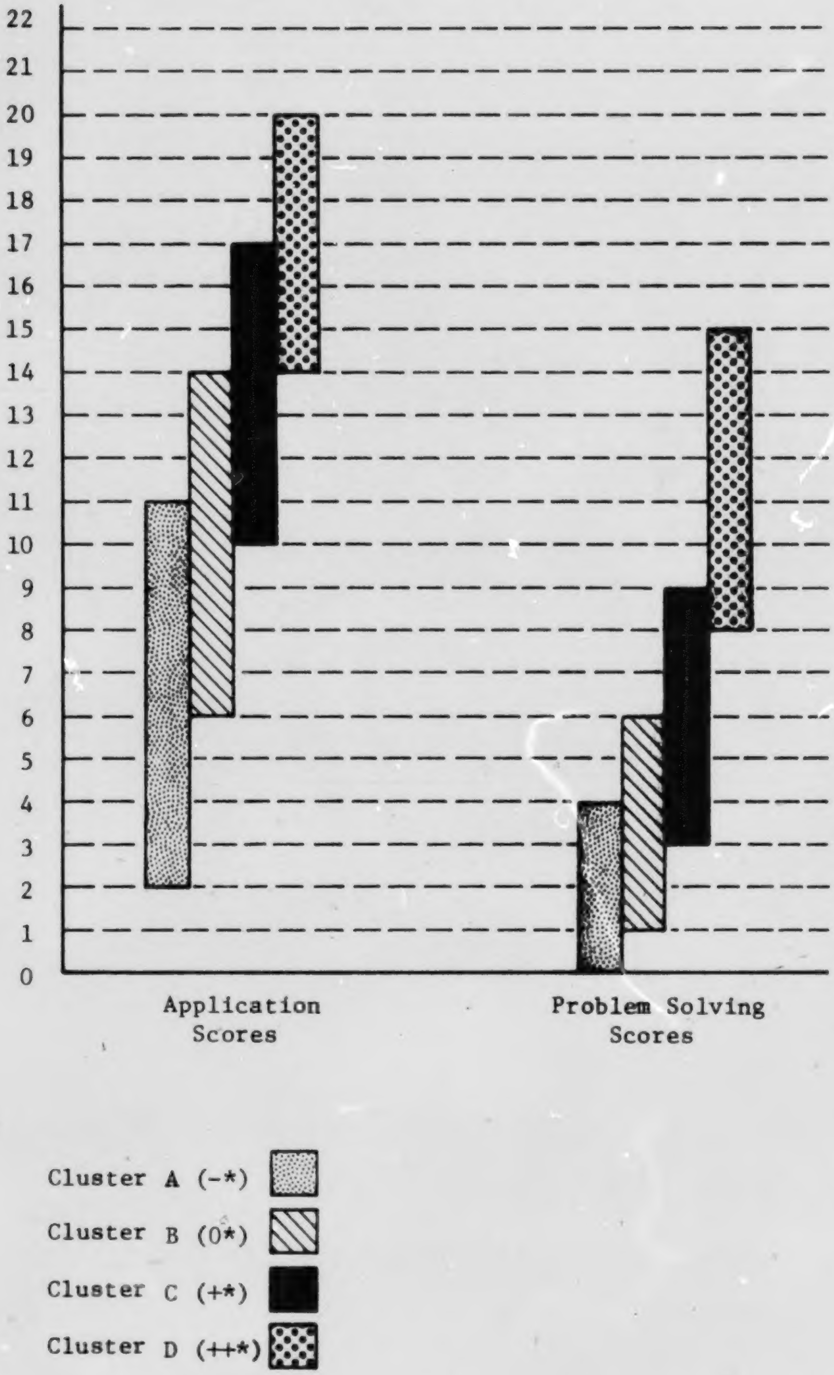


Figure 5.3 Range of the scores within each cluster.

3. All but one of the children in Cluster C was in Cluster 3, the remaining child was in Cluster 2.
4. Fifteen of the 16 children in Cluster D were in Cluster 4, one child was in Cluster 3.

Therefore, it may be said that the comprehension scores had the seeming effect of placing six children in a lower category and 89 in a higher category. This is apparent if one considers the mean scores of the children in the (0) cluster. These scores were slightly below the mean scores while close in the (0*) cluster are approximately at the mean.

There is one important difference between the clusters produced using all three scores and those using only two scores; the range of the problem solving scores within a cluster is smaller when the comprehension scores are not used than when they are used, that is, the groups are "more" homogeneous with respect to the problem solving scores. In the initial clustering procedure, there were children in the (-) group, for example, who had problem solving scores above the mean. These were the children who responded correctly to fewer application and comprehension items than most children and yet at the same time responded correctly to more problem solving items than was the mean; there were four such children. The range of the problem solving scores for Clusters 1, 2, 3, and 4 was 7, 8, 8, and 7, respectively. The range for the clusterings in which the comprehension scores were not used was 3, 6, 7, and 8 for Clusters A, B, C, and D, respectively. This difference may provide slightly different results when other measures are correlated with the scores of children within a cluster.

5.5 The Complexity of the Items

An insight into the relative complexity of the three types of items is provided by Guttman (1954). Guttman's theory is concerned with studying the order of complexity of a set of variables. The order of complexity is determined by the intercorrelations of the variables. The intercorrelations in a perfect simplex are generated by the law $r_{jk} = a_j/a_k$ where r_{jk} is the correlation between the jth and kth variables and a_j and a_k are the simplex loadings for these variables. If the intercorrelations do form a perfect simplex, the variables are said to have a simple order of complexity.

The correlation between the comprehension and the application items was .641, it was .667 between the application and problem solving items; and it was .431 between the comprehension and the problem solving items. The matrix of intercorrelations is shown in Table 5.9.

TABLE 5.9

INTERCORRELATIONS OF THE ITEM

	C	A	P
C	1		
A	.641	1	
P	.431	.667	1

The intercorrelations "almost" form a perfect simplex; to have been a perfect simplex the correlation between the comprehension and the problem solving scores would have been .428 rather than .431. Consequently, the problem solving items are more complex than the application items which, in turn, are more complex than the comprehension items. This relationship would appear to give further support to the hypothesis that the items on the test fit their respective definitions.

It has been noted previously that the items are in order of difficulty; however, there is a difference between more difficult and more complex. Difficulty in test theory connotes a relationship between group means. Consequently, it is always possible to structure tests to make one type of behavior more difficult than another. For example, if one hypothesized that it is more difficult to integrate a function than to differentiate a function, one need only administer two tests of n items to a group of people and then select the q most difficult integration items for one test and the q easiest differentiation items for another test to have the results support the hypothesis. Complexity, however, is defined in terms of correlation coefficients and a correlation coefficient is invariant under any linear transformation of scores. Hence, changing group means need not change the correlation coefficients, in fact, the intercorrelations with other tests in a simplex can be essentially the same even if the order of difficulties is reversed. The reason for this is that the correlation coefficients are dependent upon the rank order of the people taking the test, the scores may change but the rank order of the subjects does not change. Therefore, Pearsonian coefficients usually do not vary much since they are closely related to rank correlations.

Thus, one may state that the items which comprise the superitems are not only increasing in difficulty; they are also increasing in complexity.

5.6 Summary

Section 5.2 contained a discussion of an item analysis. All but five of the items on the test had a β or biserial correlation value which was considered to be either acceptable or marginally acceptable.

A clustering procedure was described in Section 5.3. The clustering procedure appeared to produce four clusters of children; one group had mean scores which were below the mean, one at the mean, one above the mean, and one considerably above the mean scores for the entire group. The results of a second clustering were described in Section 5.4. It was found that the clusters produced by deleting the comprehension scores were similar to those formed by considering all three scores with the exception that the range of the problem solving scores within a cluster was smaller when the comprehension scores were not used to form the clusters.

An inspection of the intercorrelations of the items indicated that the items were in order of complexity; this was discussed in Section 5.5.

Chapter VI

CONCLUSIONS

6.1 Introduction

This final chapter of the thesis includes a precis of the study, a discussion of the limitations of the study, and an indication of possible directions for future research. In particular, Section 6.2 summarizes the study, a causerie of the limitations is contained in Section 6.3, and preliminary results of the other studies using the test are reported in Section 6.4. The significance of the study is considered in Section 6.5 and Section 6.6 contains the implications for future research in mathematical problem solving.

6.2 Summary and Conclusions of the Study

The purpose of the study was to develop a test of mathematical problem solving behavior which provided information about the child's mastery of the prerequisites of the problem solving questions. To provide this additional information about the child's mastery of the prerequisites, each problem solving question was preceded by two other questions. One question assessed the child's understanding of the information contained in the item stem of the problem solving questions and a second question assessed the child's knowledge of an underlying concept of the problem solving question; the first question was referred to as the comprehension question and the second, the application question. The study was concerned with (1) determining if asking multiple questions on the same unit of information affects the response to those questions, and (2) measuring the extent to which the questions preceding the problem solving question were assessing prerequisite behaviors of the problem solving question.

To determine if the response to the items was affected by the other items within the superitem, six tests were constructed consisting of subsets of the items. Three of the tests contained only one type of item, comprehension, application, or problem solving, and three other tests contained two types of items. Means for each of the three scales, Comprehension, Application, and Problem Solving, were compared across the four instruments containing the scale. To determine if the differences were significant, t-tests were computed. The decision to use t-tests rather than an overall test of significance was prompted by a concern for Type II errors; it was felt that a procedure which was liberal in identifying significant differences would be appropriate when one was concerned for errors of this type. Five of the 18 t-tests computed were significant at the .05 level, two of which also were significant at the .01 level. One of the significant differences involved the Comprehension scale, three the Application scale, and one the Problem Solving scale.

Tow hypotheses were advanced to account for these significant differences: (1) the children did not have the same amount of time to respond to a particular group of items on all of the tests containing that group of items, and (2) asking multiple questions on the same unit of information affects the response to the questions. The conclusion reached was that four of the five significant differences were probably the result of the administration times of the tests, that is, that the children did not have the same amount of time to respond to a particular category of items on each test containing those items. It was felt the remaining significant difference was the result of one category of items, comprehension, having a debilitating effect upon another category, problem solving, when the two types of items appeared together on the same instrument without the inclusion of the third category of items, application.

Conditional probabilities were computed for each superitem in an effort to ascertain if the comprehension item was assessing a prerequisite of the application item and if both comprehension and application items were assessing prerequisites of the problem solving item. The conditional probabilities computed were as follows:

Prob(comprehension item correct | application item correct)

Prob(comprehension item correct | problem solving item correct)

Prob(application item correct | problem solving item correct)

Prob(comprehension and application items correct | problem solving item correct)

If all four of the conditional probabiliites computed for a given superitem were at least .75, the item was considered to be acceptable. A superitem was considered to be marginally acceptable if all four of the probabilities were at least .50. If one of the four conditional probabilities were at least .50, then the superitem was placed in the Unacceptable category; five of the 22 superitems were in this category. After examining the five superitems, it was conjectured that two of these five superitems were in this category for reasons other than the superitems not fitting the test model, that is, for reasons other than the initial two items not assessing prerequisite behaviors of the problem solving item.

Reliability of the test was estimated by two procedures. A Hoyt procedure was used under the assumption that the items were independent and a generalized KR-20 in the event the items were interdependent. The Hoyt estimate was .84 and the KR-20 was .79. The unit for computing the Hoyt estimate was the item and the superitem was the unit for the KR-20 estimate. The results of examining the mean scores of each scale on the tests containing it had led to the conclusion that the items probably were independent. Therefore, the .79 estimate was considered to be a conservative estimate and the Hoyt estimate was felt to better represent the reliability of the test.

The only indicator of validity that could be obtained was the content validity of the test as measured by a panel of judges interests in research in problem solving. Six judges were given the definitions of the three categories of items and the items in a random order and were asked to classify the items. The judges' classification of the items was compared to the classification of the items on the test and a measure of association was computed between these two classifications; the computed measure of association was .78. This concludes the initial analysis of the data.

Secondary analyses of the data included an item analysis, two cluster analyses, and an examination of the relative complexity of the questions. The GITAP (Baker, 1963) item analysis indicated there were eight items, of a total of 66, whose values for β failed to meet the generally accepted level of a "good" item (Harris, 1968); three of these items had values between .25 and .30 indicating they were marginally acceptable.

A cluster analysis (Johnson, 1967) was used to discover if there was any structure inherent in the data, that is, if the data produced natural arrangements of homogeneous groups of children. The analysis appeared to produce four groups of children. One cluster consisted of children whose mean scores on the three scales were all below the mean for the entire group; this group was designated as the (-)

cluster. A second cluster, the (0) cluster, consisted of children whose mean scores on the three scales were approximately the mean of the scales. A third cluster contained children whose mean scores on the three scales were above the mean; this cluster was designated as the (+) cluster. The final cluster, the (++) cluster, consisted of children whose mean scores on the Application and Problem Solving scales was more than two standard deviations above the mean.

Conditional probabilities for the superitems were examined for each of the four clusters. The number of superitems fitting the model was fewer in the (-) cluster than in the (++) cluster. Ten superitems were acceptable or marginally acceptable in the (-) cluster and 21 in the (++) cluster. It was felt this may have been a reflection of the children in the (++) groups not responding as erratically to the items which were multiple choice as the children in the (-) group.

A second clustering was performed, this time without considering the comprehension scores. It was of interest to note if the clusters produced without considering the comprehension scores were similar to those formed using all three scores. The second clustering procedure formed groups similar to the first procedure with the exception that the range of the problem solving scores within a cluster was smaller with the second clustering procedure than with the first.

The last of the analyses was an investigation of the relative complexity of the items on the test. A Guttman (1954) analysis indicated that the three categories of items were in order of complexity, that is, that the problem solving items were more complex than the application items which were more complex than the comprehension items. Thus, the items comprising the superitems are not merely in order of difficulty but are assessing increasingly complex behavior.

6.3 Limitations of the Study

An integral part of the study was examining the effect of asking multiple questions on the same unit of information upon the response to those questions. Independence of the items was investigated by administering the three categories of items alone, in pairs, and all together. Means for each category of item were compared across all instruments containing the scale. The conclusion of this investigation was that the significant differences among the mean scores was the result of the children not having the same amount of time to respond to a given category of item on all tests containing the category; an exception was the significant difference reported on the problem solving scale which was attributed to a debilitating effect of the comprehension item when the two were assessed as a pair on a test. Obviously, a stronger argument for the independence of the items would have existed had there been no significant differences among the means. As it is, the doubt remains as to the independence of the items.

Conditional probabilities were computed in an effort to determine if the first two items in each superitem, the comprehension and application items, were assessing prerequisite behaviors of the problem solving item. The underlying hypothesis of this procedure was that a high conditional probability implies that a correct response to one item was a prerequisite for another item. However, it may be argued that the conditional probabilities merely are indicating whether or not it is likely a child who can respond correctly to one item also will respond correctly to another, not that one item is assessing a behavior needed to respond to another item. What may be concluded from the data analysis is that the comprehension, application, and problem solving items represent categories of items which are assessing increasingly more difficult and complex behaviors. It also may be concluded that the conditional probabilities serve in a negative, if not positive, capacity by indicating when superitems are not fitting the test model, that is, when one item is not assessing a prerequisite behavior of another item. Nevertheless, it would be difficult to substantiate the assertion that the comprehension and application items are assessing prerequisites of the problem solving items from the results of the analyses in this study.

Limitations of the study include limitations of the test itself. The test proved to be too long for the population of the study.

Although most children responded to items on every page of the test, the number of items omitted in the latter portion of the test was greater than in the earlier portions. The largest number of recorded omits to a problem solving item in the first 14 superitems was 58 on superitem 4; the smallest number for superitems 15 through 22 was 65 with 142 children omitting the problem solving item of superitem 20. It is possible that had the test been shorter, the children may have persisted on some of the problem solving items. This contention is supported by the results of the test containing only the problem solving items; the mean score on the P test was 4.24 whereas the mean on the CAP test was 3.30. The difference between these two means was not significant but it would seem to indicate that the children may have devoted less time to a problem on the CAP test due to the sheer volume of the test. After responding to a few of the problem solving items, many children glanced at the last page of the test, noted the test contained 15 pages, and appeared to become discouraged. Additional weight to the argument may be provided by noting the mean score on the present test of 22 superitems is approximately the same as the mean score on an earlier version of the test which contained 19 superitems, 16 of which are on the present form of the test.

A further limitation of the study is related to the labels used to categorize the items, the labels of comprehension, application, and problem solving. If the test is used in conjunction with the

DMP mathematics program and the children have progressed through the topics contained in the suggested K-4 sequence, then the labels attached to the item categories would appear to describe the behavior required by the item. However, it is possible that for some children the application items may not represent an application of a known concept to a situation but may represent problem solving situations. For example, if a child is familiar with the concept of area but had never been asked to compute the area of a rectangle, being asked to do so might constitute problem solving behavior for this child. Thus, one should be somewhat cautious in assuming the three category of items represent that description of behavior for all children.

The limitations of the study discussed in this section may provide a direction for future considerations of the test.

6.4 Studies Using the Problem Solving Test

An earlier version of the test was used in a study by Meyer (1976). The test used by Meyer contained 16 of the 22 superitems of the test in its present form. Meyer was interested in the relationship of mathematical problem solving performance and intellectual abilities of fourth-grade children. Factor analytic and canonical correlation analysis identified the application score as being primarily of a conceptual nature. F-ratios and stepwise discriminant analysis procedures showed that the cognitive structure of those with high scores on all three scales was different than those who had high scores on the Comprehension and Application scales but low on the Problem Solving scale. (A score was considered to be high on the Comprehension and Application scale if it was above the mean; a score on the Problem Solving scale was considered high if it was at least seven out of a possible score of 19.)

A study by Whitaker (in preparation) used the test in its present form. He was considering the relationship between selected noncognitive factors and the problem solving performance of fourth-grade children. Whitaker's study included the population described in this paper and a group of children who were not using the DMP mathematics program. Preliminary results associated with the problem solving test indicate the mean comprehension and application scores of the DMP children were significantly higher than the mean scores of the non-DMP children; the difference between the problem solving scores was not significant. The means, variances, and reliabilities of each of the scales for both DMP and non-DMP children are given in Table 6.1.

The Hoyt reliability estimates indicate the test and its associated scales appear to be more reliabile with the non-DMP children than with the DMP children used in this study. An item analysis, GITAP (Baker, 1963), of the test results of the non-DMP children indicated there were nine items whose values for the item parameter β failed to meet the .30 criterion for a "good" item (Harris, 1968); four of these values were less than .25 indicating they were neither acceptable nor marginally acceptable. Two of

TABLE 6.1

MEANS, VARIANCES, AND RELIABILITY ESTIMATES OF THE SCALES FOR DMP AND NON-DMP CHILDREN

Scale	Group	Number	Mean	Variance	Reliability
C	DMP	317	15.53	9.74	.63
	NON-DMP	295	14.46	14.09	.74
A	DMP	317	10.01	13.15	.71
	NON-DMP	295	8.95	17.28	.79
P	DMP	317	3.30	6.13	.60
	NON-DMP	295	3.11	6.41	.64
CAP	DMP	317	28.84	64.18	.84
	NON-DMP	295	26.51	85.82	.89

the nine items also included in the list of "poor" items when the results were examined for the DMP children, the application item of superitem 13 and the problem solving item of superitem 20.

The test was altered to a multiple-choice format for a study investigating sex related differences in mathematics learning and related cognitive affective variables (Fennema, in preparation). The population for this study was sixth-, seventh-, and eighth-grade children. The means, variances, and reliability estimates of the scales for each grade level are given in Table 6.2.

TABLE 6.2

MEANS, VARIANCES, AND RELIABILITY ESTIMATES OF THE SCALES FOR BOYS AND GIRLS IN GRADES 6, 7, AND 8

Scale	Grade Level	Sex	Mean	Variance	Reliability
C	6	B	18.54	8.42	.71
		G	18.30	8.67	.70
	7	B	18.63	9.62	.76
		G	19.22	6.24	.66
	8	B	19.45	6.76	.72
		G	19.86	4.47	.61
A	6	B	15.43	10.69	.69
		G	14.66	14.02	.76
	7	B	16.10	13.85	.78
		G	16.01	11.58	.74
	8	B	17.58	10.45	.75
		G	17.69	8.64	.70
P	6	B	9.31	19.69	.79
		G	8.01	16.07	.75
	7	B	10.04	20.78	.79
		G	9.68	18.75	.77
	8	B	11.91	21.75	.81
		G	11.99	19.87	.79

TABLE 6.2 (Continued)

Scale	Grade Level	Sex	Mean	Variance	Reliability
CAP	6	B	43.29	89.50	.89
		G	40.97	90.56	.89
	7	B	44.77	104.85	.91
		G	44.91	85.51	.88
	8	B	48.98	85.16	.90
		G	49.54	72.28	.88

Fennema reports no significant differences between the mean scores for boys and girls on each scale at all three grade levels.

A few generalizations may be made about the test on the basis of these results.

1. The mean scores on each scale increase with the grade level of the children taking the test.
2. The comprehension scale appears to provide little information about the children in these grade levels as most children are responding correctly to the items indicating they are familiar with the vocabulary contained in the item stems. If one adopts as an arbitrary criterion of appropriateness of a test that two standard deviations

above the mean is a possible score, that is, does not exceed the total test score, then the Comprehension scale is inappropriate for all three grade levels and the Application scale is appropriate only at the sixth grade. However, the intent of the application items is to provide information about a prerequisite concept of the problem solving items, a concept with which it is hoped the children are familiar. In this context, one expects relatively high scores and yet still wishes to be alerted when a child does not possess the prerequisites of a problem solving item and, in this light, the Application scale is appropriate at all three grade levels.

3. There is a tendency for the Comprehension and Application scales to become less reliable with an increase in grade level for the girls in the study. The Comprehension scale has an estimated reliability of .70 for sixth-grade girls and .61 for eighth-grade girls. Both scales appear to be most reliable with seventh-grade boys, .76 on the Comprehension scale and .78 on the Application scale. With the exception of the results for sixth-grade boys, all of the scales tended to be the same or more reliable for boys than girls in this sample. The Problem Solving scale became more reliable with an increase in grade level for boys and girls. The overall reliability of

the problem solving test remained constant for all three grade levels.

A GITAP (Baker, 1973) item analysis on the results for the sixth-grade girls identified eight items whose values for the item parameter were less than .30; this was the larger number of items failing to meet the criterion of a good item for any of the six population groups used in the Fennema study. The number of "poor" items at the eighth grade level for both boys and girls was three. Two of the items identified by one of the item analyses for the population of the Fenneam study as a "poor" item were also included in a similar list as a result of an item analysis of the results of the fourth-grade DMP children, the comprehension item of superitem 6 and the problem solving item of superitem 19.

Another form of the test was tried out with 130 fourth-grade children using the DMP mathematics program in New York in June, 1976. This form of the test was also multiple-choice but did not contain superitems 13, 15, or 19. The results of the tryout of this version of the test are not presently available.

6.5 Significance of the Study

The study has demonstrated that the format of the problem solving test, a test consisting of superitems, is a viable form for assessing problem solving behavior. Notwithstanding the cautions outlined in Section 6.3, it appears the items within each superitem are independent, that is, the presence of other questions on the same unit of information does not affect the response to the individual items. The advantage of the format lies in the opportunity to obtain a better measure of problem solving behavior for a child by considering only the problems for which he has demonstrated some degree of preparedness by responding correctly to the first two items within a superitem. This advantage will not only prove useful to the researcher but to the classroom teacher as well.

A procedure those using the test may wish to consider is obtaining an "opportunity to learn" ratio (Fletcher, 1971), a ratio of problem solving and application scores; opportunity to learn may be a misnomer in this situation, perhaps, "opportunity to solve" would be a more apt description. Such ratios might prove to be fruitful for children with low application scores. For example, a child with an application score of eight and a problem solving score of four has solved one-half of the problems for which he has demonstrated preparedness, assuming, of course, that a correct response to a problem solving item implies a correct response to the associated application item. On the other hand, a child who has responded correctly to 12 application items and four problem solving items has solved only one-third of the problems for which he has indicated mastery of a prerequisite concept. Although the problem solving score of these two children is the same, it would appear that one child is more of a problem solver than the other. An examination of "opportunity to learn" ratios for the children in the clusters using the mean application and problem solving score for that cluster suggests that the children in the (+) and (++) clusters are not responding merely to more of the application and problem solving items, but that the ratio of problem solving to application scores is also greater; the ratios for clusters (-), (0), (+), and (++) are .252, .243, .373, and .589, respectively.

The test should prove to be an efficacious instrument in studies interested in correlating measures with problem solving behavior. The efficaciousness stems from the availability of three scores with which to correlate measures as well as the scores of groups of children formed by the clustering procedure.

6.6 Implications of Future Research

The first part of this section proposes points to consider in replicating the study and the latter portion of the section suggests avenues for future study.

Two major limitations of the study were the existence of significant differences among the mean scores of the three scales on the various tests containing them and the inability to conclude the initial two items within a superitem were assessing prerequisite behaviors of the problem solving item. A better approximation of the time necessary for the administration of the tests consisting of subsets of the items is necessary. An approximation of the necessary administration time might be obtained by administering the complete test to a group of children and noting the length of time needed to respond to each of the categories of items. The difficulty in arriving at an exact time would still exist due to the items frequently sharing an item stem and the child may or may not read the stem more than once. However, by careful observation, it will be possible to arrive at a better estimate of the necessary administration times of the tests than those used in this study. In particular, it is clear that the CA and AP tests should not share the same administration time as they did in this study.

The second weakness, the inability to say with certainty that the initial two items are assessing prerequisite behaviors of the problem solving item, may be overcome by asking a panel of judges to note specifically if these items are assessing prerequisite behaviors of the problem solving items. The doubt would still remain, but more conclusive results would be obtained.

The limitation of the test length may be overcome by constructing a test of fewer items.

The multiple-choice format of the test developed for a study by Fennema (in preparation) and subsequently used by Whitaker (in preparation) needs further study. Adjusting the test to a multiple-choice format entailed changing the format of 13 of the problem solving items; nine of the 22 items were already in that form, three of which consisted of listing all possible outcomes of an event. A similar situation existed among the application items as ten were already in a multiple-choice format, two of which consisted of listing all possible outcomes. Sixteen of the comprehension items were in a multiple-choice format on the original test, thirteen of which were true-false statements.

There are those who believe a multiple-choice format is inapppropriate for assessing problem solving behaviors (Zalewski, 1976) and others (ETS, 1963) who believe it is possible to assess behavior at all levels in the cognitive domain by a multiple-choice item. It is difficult at this juncture to determine the effect of the two forms of the test. The fourth-grade children who took the multiple-choice form of the test also took the original form, although 11 weeks apart, thereby making it difficult to arrive at any conclusions as to the effect of the format on the responses. The results which were available for the fourth-grade group indicate the application and problem solving scores were significantly higher on the multiple-choice form of the test than on the initial form (see Table 6.3). It is difficult to determine if the significant differences were the result of the test format, learning during the 11 weeks, or the result of the children taking the test for the second time. It is possible the children may have remembered, when taking the test the second time, how they had responded to an item on the test the previous time, thereby providing them with additional time to solve problems. The difference between the problem solving score on the multiple-choice format of the test was not significantly

TABLE 6.3

RESULTS OF NON-MULTIPLE CHOICE AND MULTIPLE CHOICE FORMS OF THE TEST

Scale	Test Format	Mean	Variance	Reliability	Significant Difference
C	MC	15.81	9.44	.66	
	N-MC	15.53	9.74	.63	
A	MC	11.68	14.19	.71	$p < .01$
	N-MC	10.01	13.15	.71	
P	MC	4.76	8.96	.63	$p < .01$
	N-MC	3.30	6.13	.60	
CAP	MC	32.25	70.45	.85	$p < .01$
	N-MC	28.84	64.18	.84	

different from the problem solving score on the P test; the higher problem solving score on the P test had been attributed to the children having more time to respond to the problem solving items on this test than on the other tests containing the same items.

The desirability of a multiple-choice test in situations where the test will be administered to large groups of children provides a motivation for investigating the effect of the format upon the problem solving score. It also will be of interest to note the effect of the format upon the results of clustering procedures.

The problem solving test developed in this study may be viewed as a step in a series to investigate problem solving behavior. The tests assists in determining whether or not a child possesses the requisite skills and level of concept mastery for the problem solving item. Although this is not separating concept attainment from the problem, it is providing an indication when the child could not be expected to solve a problem. A stronger instrument would be one in which computational errors did not affect the problem solving score.

Another step in the series to investigate problem solving behavior may be to gain insight into how the child is solving the problem, to obtain a measure of the procedures or processes used by the child in solving the problem. Investigators have attempted to elicit this information by envelope or tab-item tasks. The difficulty with this approach is that it appears to lend itself to the investigation of problems which are essentially arithmetic in nature; a counter example is the task developed by Neimark and Lewis (1967) described in Chapter II.

It also may prove interesting to examine the pattern of responses to the problem solving items on the test. The test contains a variety of problem solving situations; consequently, it is possible to consider types of situations to which a child has responded correctly. For example, it may be informative to consider the correct responses of children who responded correctly to the cube rotation item, to note if any pattern emerges in the types of problems solved by these children.

In conclusion, the results of the study support the contention that a test composed of superitems is a viable form for assessing problem solving behavior. The test should prove to be an effective instrument in studies wishing to correlate measures with problem solving behavior. The effectiveness is derived from the existence of three scores produced by the test with which to correlate measures as well as the scores of groups of children formed by the clustering procedure. It is hoped the test or one of the avenues of research suggested in this section will serve as a source of illumination of that ellusive human behavior called problem solving.

REFERENCES

American Psychological Association. _Standards for educational and psychological tests and manuals_. Washington D.C.: Authors, 1966

Alluisi, E.A., & Coates, G.D. Parameters and factor structures of a three-phase code transformation task (3P-COTRAN). _Perceptual and Motor Skills_, 1969, 29, 155-166.

Asher, J. J. Towards a neo-field theory of problem solving. _Journal of General Psychology_, 1963, 68, 3-8.

Baker, Frank B. Generalized item and test analysis program (GITAP). _Educational and Psychological Measurement_, 1963, 23, 187-190.

Bassler, Otto C.; Beers, Morris J.; & Richardson, Lloyd I. Comparison of two instructional strategies for teaching the solution of verbal problems. _Journal for Research in Mathematics Education_, 1975, 3, 170-177.

Bourne, Lyle G.; & Battig, William F. Complex processes. In J. Sidowski (Ed.) _Experimental methods and instrumention in psychology_. New York: McGraw-Hill, 1966, p 541-576.

Branca, Nicolas. _Strategies in learning mathematical structures: A brief report of the study_. April, 1971. Eric Document Number ED 049 056.

Burch, R.L. Formal analysis as a problem solving procedure. _Journal of Education_, 1953, 136, 44-47.

Buswell, Guy T. Patterns of thinking in solving problems. _University of California Publications in Education_, 1956, 12, 63-148.

Campbell, Clinton I. Blind variations and selective retention in creative thought as in other knowledge processes. _Psychological Review_, 1960, 67, 380-400.

Chase, Clinton I. The position of certain variables in the prediction of problem solving in arithemtic. _Journal of Educational Research_, 1960, 54, 9-14.

Coates, G.D.; Alluisi, E.A.; & Morgan, B.B. Trends in problem solving research: Twelve recently described tasks. _Perceptual and Motor Skills_, 1971, 33, 495-505.

Cochran, W.G., & Cox, G.M. _Experimental designs_. New York: John Wiley & Sons, 1957.

College Entrance Examination Board Commission on Mathematics. Program for college preparatory mathematics. Report. New York: College Entrance Examination Board, 1954.

Covington, J.; Crutchfield, R.; & Davies, L. The productive thinking program, series I. General problem solving. Berkeley: Breselten Printing Company, 1966.

Cureton, Edward E. Reliability and validity: Basic assumptions and experimental designs. Educational and Psychological Measurement, 1965, 25, 327-346.

Dahmus, J.E. How to teach verbal problems. School Science and Mathematics, 1970, 10, 121-138.

Davis, G.A. Current studies of research and theory in human problem solving. Psychological Bulletin, 1966, 66, 36-54.

Duncan, C.P. Effects of instruction and information on problem solving. Journal of Experimental Psychology, 1963, 65, 321-327.

Duncker, Karl. On problem sovling. Psychological Monograph, 1945, 58, 1-111.

Durost, W.N.; Bixler, H.H.; Wrightstone, J.W.; Prescott, G.A.; & Balow, I.H. Metropolitan achievement tests: Advanced form F. New York: Harcourt, Brace, Javanovich, 1971

Dye, Newal W.; & Very, Philip S. Growth changes in factorial structure by age and sex. Genetic Psychology Monographs, 1968, 78, 55-88.

Educational Services Incorporated. Goals for school mathematics. Report of the Cambridge Conference on School Mathematics. Boston: Houghton Mifflin Co., 1963

Educational Testing Service. Multiple-choice questions: A close look. Princeton: Authors, 1963.

Egan, D.E.; & Green, J.G. Acquiring cognitive structure by discovery and rule learning. Journal of Educational Psychology, 1973, 64, 85-97.

Fattu, N.A.; & Mech, E.V. The effect of set on performance in a "trouble shooting" situation. Journal of Psychology, 1953, 36, 153-163.

Fehr, Howard F.; & Phillips, J.M. Teaching modern mathematics in the elementary school. Redding, Massachusetts: Addison-Wesley, 1967.

Fennema, Elizabeth. Sex related differences in mathematics learning and related cognitive affective variables. In preparation.

Fletcher, Harold J. An efficiency reanalysis of the results. Journal for Research in Mathematics Education, 1971, 2, 143-156.

French, John W. The relationship of problem solving style to the factor composition of tests. Educational and Psychological Measurement, 1965, 25, 9-28.

French, R.S. The effect of instructions on the length-difficulty relationship for a task involving sequential dependency. Journal of Experimental Psychology, 1954, 48, 89-97.

Glaser, R.; Damrin, D.; & Gardner, G. The tab item: A technique for the measurement of proficiency in diagnostic problem solving tasks. Educational and Psychological Measurement, 1954, 14, 283-293.

Guildford, J.P. The structure of the intellect. Psychological Bulletin, 1956, 53, 267-293.

Guttman, Louis. A new approxh to factor analysis: The radex. In P.F. Lazarfeld (Ed.) Mathematical thinking in the social sciences. Glencoe, Illinois: The Free Press, 1954.

Gyr, J.W. An investigation into, and speculation about, the formal nature of a problem solving process. Behavioral Science, 1960, 5, 39-59.

Harris, Margaret L. Some methodological suggestions for construction of an objective measurement instrument. Technical Memo Number M-1968-2. Madison: Wisconsin Research and Development Center for Cognitive Learning, 1968.

Hays, William L. Statistics for the Social Sciences. New York: Holt, Rinehart and Winston, Inc., 1973.

Hieronymus, AN.N, & Lindquist, E.F. Iowa test of basic skills, Form 6. Boston: Houghton Mifflin Company, 1971

Hilbert, David. Mathematical Problems. In Archives of Mathematics and Physics. 1906

Hoyt, C. Test reliability estimated by analysis of variance. Psychometrika, 1941, 12, 153-160.

Immerzeel, G.; Wiederanders, D.; & Scott, L. Ginn elementary mathematics. Lexington, Massachusetts: Ginn and Company, 1972.

Jerman, Max. Problem solving in arithmetic as transfer from a productive thinking program. Unpublished doctoral dissertation, Stanford University, 1971.

Jerman, Max. Individualized instruction in problem solving in elementary school mathematics. Journal for Research in Mathematics Education, 1973, 4, 6-19.

John, E.R.; & Miller, J.G. The acquisition and application of information in the problem-solving process: An electronically operated logical test. Behavioral Science, 1957, 2, 291-301

Johnson, E.S. An information-processing model of one kind of problem-solving. Psychological Monograph, 1964, 78, No.4

Johnson, Harry C. Problem-solving in arithmetic: A review of the literature. Elementary School Journal, 1944, 44, 396-403, 476-482.

Johnson, Stephen C. Hierarchical clustering schemes. Psychometrika, 1967, 32, 241-254.

Katona, G. Organizing and memorizing. New York: Columbia University Press, 1940.

Kelley, Truman; Madden, Richard; Gardner, Eric;& Rudman, Herbert. Directions for administering intermediate I battery. New York: Harcourt Brace & World, Inc., 1964.

Kendler, H.H.; & Kendler, T.S. Vertical and horizontal processes in problem solving. Psychological Review, 1962, 69, 1-16.

Kilpatrick, Jeremy. Analyzing the solution to word problems in mathematics: An exploratory study. (Doctoral dissertation, Stanford University) Ann Arbor, Michigan: University Microfilm, 1967, No. 68-6442.

Kline, Morris. Why Johnny can't add: The failure of the new math. New York: Vintage Books, 1973.

Laughery, K.R.; & Gregg, I.W. Simulation of human problem solving behavior. Psychometrika, 1962, 27, 265-282.

Lazerte, J.E. The development of problem solving abiltiy in arithmetic. Toronto: Clarke Irwin, 1933.

Loevinger, Jane. The attenuation paradox in test theory. Psychological Bulletin, 1954, 51, 493-504.

Lord, Frederic M.; & Novick, Melvin R. Statistical theory of mental test scores. Reading, Massachusetts: Addison Wesley, 1974.

Lucas, John F. An exploratory study of the diagnostic teaching of heuristic problem solving strategies in calculus. Unpublished doctoral dissertation, University of Wisconsin, 1972.

Luchins, Abraham S.; & Luchins, Edith H. New experimental attempts at preventing mechanization in problem solving. Journal of Genetic Psychology, 1950, 42, 279-297.

Mandler, G. From association to strucutre. Psychological Review, 1962, 69, 415-427.

Maier, N.R.F. Reasoning in humans: Induction. Journal of Comparative Psychology, 1930, 10, 114-143.

McGuire, Chirstine, & Babbott, D. Simulation technique in the measurement of problem solving skills: A technological or concept break-through. Paper read at American Educational Research Association National Council of Measurement Education, Chicago, February, 1965.

Means, Gladys H., & Lorel, M.Ray. The influence of problem solving ability of following an efficient model. Journal of Educational Research, 1968, 62, 135-141.

Merrifield, P.R.; Guildford, J.P., Christensen, P.R., & Frick, J.W. The role of intellectual factors in problem solving. Psychological Monographs, 1962, 76, No. 10.

Meyer, Ruth Ann. A study of the relationship of mathematical problem solving performance and intellectual abilities of fourth grade children. Technical Report No. 379. Madison: Wisconsin Research and Development Center for Cognitive Learning, 1973.

Moore, Omar Khayyam; Anderson, Scarvia B. Modern logic and tasks for experiments on problem solving behavior. Journal of Psychology, 1954, 38, 151-160.

National Advisory Committee on Mathematical Education. Overview and analysis of school mathematics K-12. Washington, D.C.: Authors, 1975.

Neimark, Edith D.; & Lewis, Nan. The development of logical problem solving strategies. Child Development, 1967, 38, 107-117.

Nelson, L. Doyal; Kirkpatrick, Joan. Problem solving. In J. Payne (Ed.) Mathematics learning in early childhood, Thirty-seventh yearbook national council of teachers of mathematics. Reston, Virginia: National Council of Teachers of Mathematics, Inc., 1975.

Newell, A.; Shaw, J.C.; & Simon, H.A. Elements of a theory of human problem solving. Psychological Review, 1958, 65, 151-167.

Norman, P.B. Relationship between problem-solving ability, computational skill, intelligence, and amount of training in mathematics. Doctoral disseration, Columbia University, 1950.

The Ontario Institute for Studies in Education. K-13 mathematics, some non-geometric aspects. Part II: Computing, logic, and problem solving. Toronto: Author, 1971.

Pereira, Waldecyr. The relationship between girls' overt verbalization, performance, retention, rules, and strategies as they learn a mathematical structure: A study based on elements of a potential theory which relates thinking, language, and learning. Technical Report No. 293. Madison: Wisconsin Research and Developmetn Center for Cognitive Learning, 1973.

Polya, G. Mathematical discovery. Vol. I. New York: Wiley, 1962, 2 vols.

Posner, J.I. Memory and thought in human intellectual performance. British Journal of Psychology, 1965, 56, 197-215.

Post, T.R. The effects of the presentation of a structure of the problem solving process upon problem solving ability in seventh grade mathematics. (Doctoral dissertations, Indiana University) Ann Arbor, Michigan, University Microfilms, 1968, No. 68-4746.

Post, Rhomas R.; & Brennan, Michael L. An experimental study of the effectiveness of a formal versus an informal presentation of a general heuristic process on problem solving in tenth-grade geometry. Journal for Research in Mathematics Education, 1976, 7, 59-64.

Ray, W.S. Complex tasks for use in human problem solving research. Psychological Bulletin, 1955, 52, 134-149.

Rees, H.J.; & Israel, H.E. An investigation of the establishment and operation of mental sets. Psychological Monograph, 1935, 46 (6, Whole No. 210).

Rimoldi, H.J.A. A technique for the study of problem solving. Educational and Psychological Measurement, 1955, 15, 450-461.

Rimoldi, H.J.A; Aghi, M.; & Burger, G. Some effects of logical structure, language, and age in problem solving in chidlren. Journal of Genetic Psychology, 1968, 112, 127-143.

Romberg, Thomas A.; & Harvey, John G. Developing mathematical processes: Background and projections. Working Paper No. 14. Madison: Wisconsin Research and Development Center for Cognitive Learning, 1969.

Romberg, Thomas A.; Harvey, John G.; & Moser, James M. Developing mathematical processes. Chicago: Rand McNally, 1974, 1975, 1976.

Romberg, Thomas A. Specifications for terminal accountability tests for DMP. Project Paper No. 74-1. Madison: Wisconsin Research and Development Center, 1974.

Saustad, P.; & Raaheim, K. Problem solving, past experience, and availability of functions. British Journal of Psychology, 1960, 51, 97-104.

Science Research Associates, Inc. Mothern math understanding test form C, multilvel edition. Chicago: Author, 1966.

Sheerer, M. Problem solving. Scientific American, 1963, 208, 118-128.

Simon, Herbert A.; Newell, Allen. Human problem solving: The state of the theory in 1970. American Psychologist, 1970, 26, 145-159.

Simon, Herbert A.; Newell, Allen. Human problem solving. Englewood Cliffs, New Jersey: Prentice-Hall, 1971.

Staats, W.S., & Staats, C.K. Complex human behavior. New York: Holt, Rinehart, and Winston, 1963.

Suppes, P.; Smith, S.; Easterday, K.; Firr, D.; Kaplan, J.; Phillips, G.; & Bredehorn, G. Random House mathematics program. New York: Random House, 1972.

Treacy, J.P. The relationship of reading skills to the ability to solve arithmetic problems. Journal of Educational Research, 1944, 38, 86-95.

Vanderlene, L.F. Does the study of quantitative vocabulary improve problem solving? Elementary School Journal, 1964, 65, 143-152.

Wertheimer, Max. Productive thinking. New York: Harper and Row, 1945.

Whitaker, Donald R. A study of the relationships between selected noncognitive factors and the problem solving performance of fourth grade children. In preparation.

Zalewski, Donald L. *An exploratory study to compare two performance measures: An interview-coding scheme of mathematical problem solving and a written test*. Technical Report No. 306. Madison: Wisconsin Research and Developmetn Ceter for Cognitive Learning, 1974.

APPENDIX A

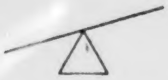
PROBLEM SOLVING TEST

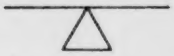
© 1975 - The Regents of the University of Wisconsin System for the Wisconsin
Research and Development Center for Cognitive Learning.
Copyright is claimed only during the period of development, test, and evaluation, unless authorization is received from the National Institute of Education to claim copyright on the final materials. For the current copyright status, contact either the copyright proprietor or the National Institute of Education.

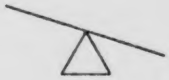
Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the National Institute of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by that agency should be inferred.

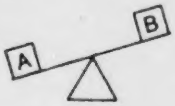
Center Contract No. NE-C-00-3-0065

Circle the picture that shows the balance beam in balance.









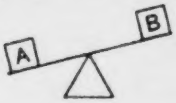
Circle the sentence that tells about the weights $\boxed{A}$ and $\boxed{B}$.

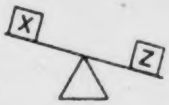
$\boxed{A}$ weighs less than $\boxed{B}$.

$\boxed{A}$ weighs more than $\boxed{B}$.

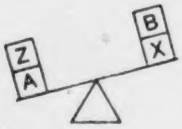
$\boxed{A}$ weighs the same as $\boxed{B}$.

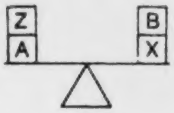
impossible to tell from the picture

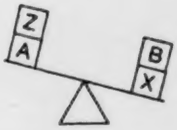




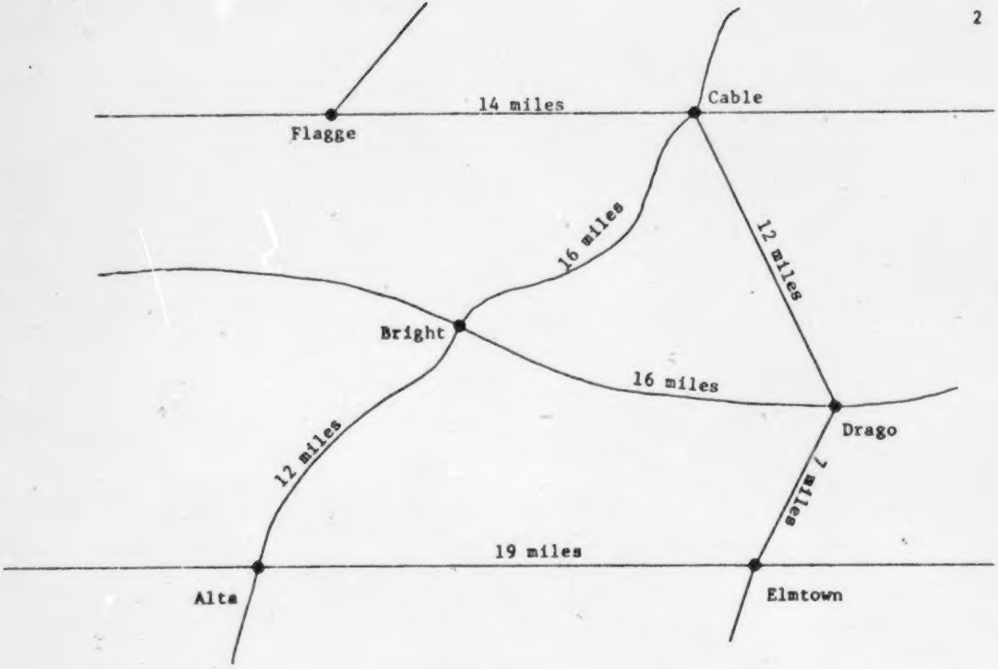
Weights $\boxed{A}$ and $\boxed{Z}$ are put together on one end of the balance beam and $\boxed{B}$ and $\boxed{X}$ are put together on the other end of the balance beam. Circle the picture that shows how the balance beam might look.







impossible
to tell


14 miles
Cable
Flagge
16 miles
12 miles
Bright
16 miles
Drago
12 miles
7 miles
19 miles
Alta
Elmtown

The distance from Alta to Bright is:
- 7 miles
- 12 miles
- 16 miles
- 19 miles

The shortest distance from Alta to Drago is:
- through Bright
- through Cable
- through Elmtown
- through Flagge

The sign

BRIGHT	16
ELMTOWN	19

should be placed:
- in Drago
- in Alta
- in Flagge
- in Cable

Jack and Bob went fishing. They both caught two kinds of fish, trout and pike. Bob caught 10 trout. Together they caught 18 trout and 10 pike. Jack caught 15 fish.

How many trout did Bob catch? ______________

How many fish did they catch altogether? ______________

How many pike did Bob catch? ______________

There are 45 houses on Century Avenue. The houses are white, blue, yellow, green, or gray. 18 of the houses are white. There are half as many blue houses as white houses. 15 of the houses are yellow. Only 1 house is gray.

There are more blue houses than white houses. TRUE FALSE Impossible to tell

There are more blue houses than yellow houses. TRUE FALSE Impossible to tell

How many of the houses are green? ______________

4

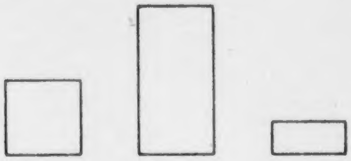
Put a circle around the sentence that tells about A < B.

A is less than B.

A is greater than B.

A is equal to B.

The sentences tell about the figures. Decide which figure is A, which is B, and which is C. Write the letters on the figures.



A < B   B < C

The following sentences tell about the numbers A, B, C, and D.

$$C + 3 = B$$

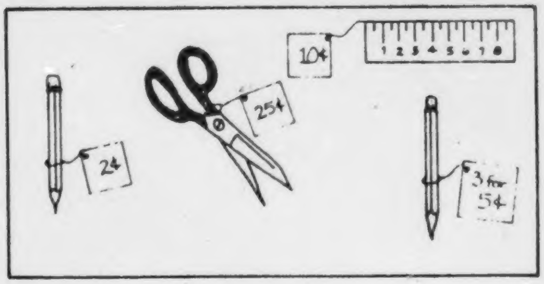
$$D < A$$

$$B = D - 4$$

Write the four numbers A, B, C, and D in order from smallest to largest.

________, ________, ________, ________
(smallest)

5



Both pencils cost the same. TRUE FALSE IMPOSSIBLE TO TELL

How much would 8 of the 2¢ pencils cost? ______________

What is the greatest number of pencils you could buy for 27¢? ______________

Jack lost some weight. Susan lost half as much weight as Jack lost. Carol lost half as much weight as Susan lost. Carol lost 10 pounds.

Jack lost more weight than Susan. TURE FALSE IMPOSSIBLE TO TELL

Circle the number that tells how many pounds Susan lost.

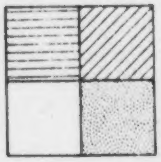
$2\frac{1}{2}$ 5 10 20 40 none of these

If Jack weighs 130 pounds now, how much did he weigh before he lost weight?

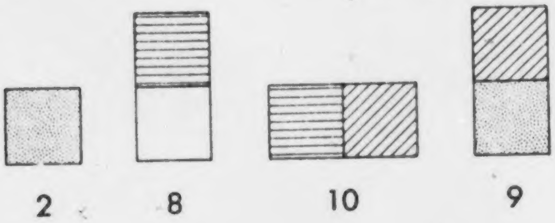
____

6

Each of the four blocks ▤, ▨, □, ░ in the picture is covering a number of chips.

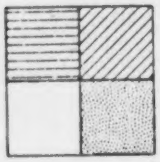


The numbers below the pictures tell how many chips are covered by those blocks.



How many chips are covered by ░ ? ____________

How many chips are covered by



? ____________

How many chips are covered by □ ? ____________

7

Jackie has twice as many football cards as Mike. Each has more football cards than Terry. Mike has 15 football cards.

Terry has the most football cards. TRUE FALSE

How many football cards does Jackie have? ______________

Circle the number that tells how many football cards the three could have altogether. (You don't know exactly how many they have altogether.)

63 42 58 72

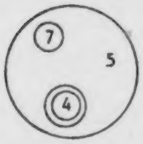
A parking lot has room for 8 rows of cars with 9 cars parked in each of those rows.

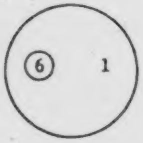
The parking lot has the same
number of cars in each of the 8 rows. TRUE FALSE

How many cars can be parked in the parking lot? ______________

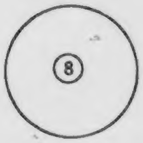
In another parking lot, trucks are parked. Each truck takes the space of 3 cars. There are 12 trucks in the parking lot and it is completely full. If there were 4 rows in the parking lot, how many cars could be parked in each row? ______________

8

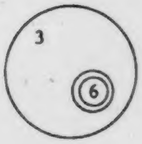
In Circleland, people write 

when they mean 475 and they write 

when they mean 61.

In Circleland, people write 

when they mean 8. TRUE FALSE

What do they mean when they write 

?

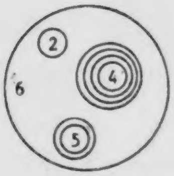
36

63

630

603

306

What do they mean when they write 

?

4,526

40,526

4,562

45,620

45,260

Everything weighs 6 times as much on Earth as it does on the Moon.

Everything weighs 2 times as much on Earth as it does on Mars.

A box weighs more on the Moon than it does on Earth.

TRUE FALSE IMPOSSIBLE TO TELL

A box weighs 24 pounds on the Moon. How much does it weigh on Earth?

$\frac{1}{6}$ 4 6 24 144 impossible to tell

A dog called Rover weighs 4 pounds on the Moon.

Another dog called Spot weighs 12 pounds on Mars.

If both dogs were in the same place, then

Rover would weigh more than Spot

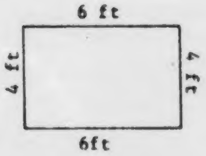
Rover would weigh less than Spot

Rover would weigh the same as Spot

Impossible to tell

10

This is a rectangle.



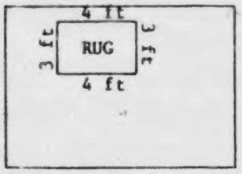
The rectangle is 4 feet long on one side and 6 feet long on another side.

TRUE FALSE

The area of the rectangle is: 10 sq ft 20 sq ft 24 sq ft 100 sq ft

The area of the floor of a room pictured below is 96 sq ft. A small rug is on the floor.

The area of the floor not covered by the rug is:



79 sq ft

82 sq ft

84 sq ft

89 sq ft

impossible to tell

A gallon is larger than a pint.

TRUE FALSE

Solve the sentence.

16 quarts = ☐ gallons

Solve the sentence.

2 gallons = ☐ cups

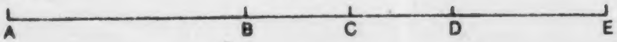
Measurements

2 cups = 1 pint

2 pints = 1 quart

4 quarts = 1 gallon

2 half gallons = 1 gallon



The distance from A to E is 23 cm.

The distance from D to A is 17 cm.

The distance from B to E is 14 cm.

C is half-way between B and D.

What is the distance from A to D? ______________

What is the distance from A to B? ______________

What is the distance from B to C? ______________

The dining room of a ship has 14 tables. One-half of the tables have 6 chairs at each table and the other half of the tables have 4 chairs at each table.

7 tables have 6 chairs at each table.   TRUE   FALSE

What is the largest number of people that can sit in the dining room at one time? ______________

Another ship has 70 passengers. If all of the passengers and crewmen are sitting in the lifeboats, there are 7 passengers and 2 crewmen in each lifeboat.

How many crewmen are on the ship? ______________

Put a number in each ☐ to complete the number pattern.

1 2 3 ☐ 5 6

4 7 10 ☐ 16 19

1 2 4 7 ☐ 16 22

Rosemarie had 30 pencils. She lost 16 of them.

She has fewer pencils now. TRUE FALSE IMPOSSIBLE TO TELL

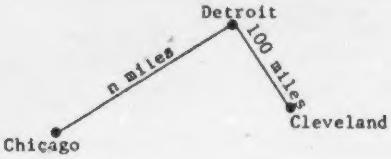
How many pencils does she have now? ______________

Kirk had some pieces of candy. He gave one-half of them to Mary.

Then he gave 3 pieces to Harold and had 6 pieces left.

How many pieces did he start with? ______________

n is the number of miles from Chicago to Detroit.



Circle the true sentence:

It takes n hours to drive from Chicago to Detroit.

n is the number of miles from Detroit to Chicago.

It takes n hours to fly from Chicago to Detroit.

n is the distance in miles from Detroit to Cleveland.

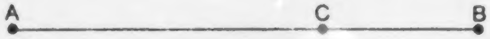
2(n) + 100 represents:

The distance from Chicago to Cleveland.

A distance twice the distance from Chicago to Detroit.

A distance less than the distance between Chicago and Detroit.

A distance more than twice the distance between Chicago and Detroit.



The distance from A to C is twice the distance from C to B.

If you drive from A to B and then back to C, you will have gone:

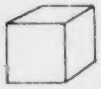
4 times the distance from A to C

2 times the distance from A to C

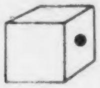
$1\frac{1}{3}$ times the distance from A to C

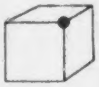
$1\frac{1}{2}$ times the distance from A to C

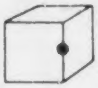
This is a picture of a cube.



Circle the picture that has a ● on a face of the cube.

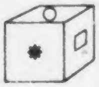


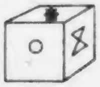


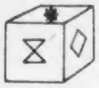


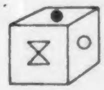
How many faces does a cube have? ______________

These are four views of the same cube.

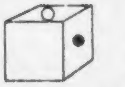








Here is another view of the same cube.



Circle the figure that goes on the □.

✱ ○ □ ⧗ ◇ ●

A ferryboat never crosses the river unless it is full. The ferryboat is full when it holds 12 cars. The ferry is also full when it holds 8 trucks. Cars and trucks are never on the ferryboat at the same time.

Sometimes the ferryboat crosses the river with 10 cars.

TRUE FALSE IMPOSSIBLE TO TELL

In two trips, how many trucks could the ferryboat have carried?

8 12 16 24 none of these

The ferryboat made 4 trips across the river and carried 44 vehicles. (Cars and trucks are vehicles)

Circle the sentence which could be true.

The ferryboat was filled with cars each time.

The ferryboat was filled with trucks each time.

The ferryboat was filled with cars more than one-half the time.

The ferryboat was filled with trucks at least one-half the time.

None of the sentences is true.

Find the sums.

$$\begin{array}{r} 6 \\ +\ 9 \\ \hline \end{array} \qquad \begin{array}{r} 463 \\ +\ 296 \\ \hline \end{array}$$

Find the missing numbers.

$$\begin{array}{r} 3\square 5 \\ +\ 25\square \\ \hline \square 43 \end{array}$$

APPENDIX B

ITEM DIFFICULTY AND X_{50}

	ITEM DIFFICULTY		
	C	A	P
1	.98	.79	.34
2	.96	.79	.24
3	.96	.57	.15
4	.67	.50	.11
5	.87	.67	.11
6	.80	.85	.11
7	.55	.30	.15
8	.86	.32	.03
9	.89	.65	.03
10	.57	.43	.05
11	.35	.24	.11
12	.66	.41	.22
13	.75	.05	.04
14	.89	.49	.11
15	.74	.08	.07
16	.48	.09	.05
17	.81	.66	.19
18	.71	.54	.20
19	.43	.16	.25
20	.63	.45	.04
21	.50	.47	.12
22	.56	.51	.16

	X_{50}		
	C	A	P
1	-51.48	-2.44	.99
2	-5.50	-1.68	1.22
3	-5.50	-.33	1.76
4	-1.58	.03	1.53
5	-3.78	-.95	2.19
6	-3.53	-2.40	2.31
7	- .26	.80	1.36
8	-3.80	1.03	2.15
9	-3.34	-.81	2.15
10	- .44	.28	4.60
11	1.39	1.56	2.28
12	-1.60	.52	1.90
13	-1.73	3.66	4.77
14	-1.92	.05	1.94
15	-1.26	2.63	2.94
16	.09	2.50	2.09
17	-1.10	-.59	1.67
18	- .73	-.13	1.34
19	.27	2.86	3.30
20	- .45	.25	8.41
21	.01	.13	.35
22	- .24	-.05	1.65

APPENDIX C

DIAMETERS OF CLUSTERINGS

CYCLE	NUMBER OF CLUSTERS	DIAMETER
302	15	3.89
303	14	4.30
304	13	4.75
305	12	4.84
306	11	5.49
307	10	5.87
308	9	8.69
309	8	11.34
310	7	11.78
311	6	17.59
312	5	19.30
313	4	25.18
314	3	49.84
315	2	80.98
316	1	294.01

APPENDIX D

K-4 TOPICS OF THE
DEVELOPING MATHEMATICAL PROCESSES
MATHEMATICS PROGRAM

TOPIC 1

Describing and Classifying

The children describe and classify familiar objects. Likenesses and differences are stressed and the skill of sorting is introduced.

REGULAR OBJECTIVES

1. Given an attribute, chooses an object that has that attribute.
2. Given two objects and an attribute, states whether the objects are the same or different on that attribute.
3. Given an object, describes it in terms of its attributes.

PREPARATORY OBJECTIVES

4. Given several objects and an attribute, sorts the objects into sets on that attribute.
5. Given a set, states an attribute that all of its members have in common.

TOPIC 2

Comparing and Ordering on Length

Real objects are directly compared and then ordered on length. The children determine whether two objects are the same length and if not, they determine which is longer.

REGULAR OBJECTIVES

1. Given two objects, compares them on the attribute of length.
2. Given two objects, orders them on the attribute of length.

PREPARATORY OBJECTIVES

3. Given several objects and an attribute, sorts the objects into sets on that attribute.
4. Given a set, states an attribute that all of its members have in common.

TOPIC 3

Equalizing on Length

After comparing and ordering two objects or pictures of objects on length, the children make them equal in length by adding on to the shorter or taking away from the longer.

REGULAR OBJECTIVE

1. Given two objects, equalizes them on the attribute of length.

PREPARATORY OBJECTIVES

2. Given several objects and an attribute, sorts the objects into sets on that attribute.
3. Given a set, states an attribute that all of its members have in common.

TOPIC 4

Ordering More than Two Objects on Length

The children put more than two objects in order from shortest to longest or from longest to shortest.

REGULAR OBJECTIVE

1. Given more than two objects, orders them on the basis of the attribute of length from longest to shortest or shortest to longest.

PREPARATORY OBJECTIVE

2. Given several objects and an attribute, sorts the objects into sets on that attribute.

TOPIC 5

Representing Length

In situations in which the children cannot compare and order lengths directly, they physically represent the lengths and compare and order the representations to decide about the original objects.

REGULAR OBJECTIVES

1. Given an object, physically represents a length of that object.
2. Given two lengths and a physical representation of at least one of them, compares and orders the lengths on the basis of the physical representation.

PREPARATORY OBJECTIVES

3. Given an object, pictorially represents a length of that object.
4. Given two lengths and a pictorial representation of at least one of them, compares and orders the pictorial representation.
5. Given several objects and an attribute, sorts the objects into sets on that attribute.
6. Given a set, states an attribute that all of its members have in common.

TOPIC 6

Movement and Direction

The children learn that a path is a representation of a movement in some direction. Simple direction words are used.

REGULAR OBJECTIVES

1. Given two points and a movement between them, physically represents that movement.
2. Given two points and a movement between them, pictorially represents that movement.

PREPARATORY OBJECTIVES

3. Given a path, describes it.
4. Given a set, states an attribute that all of its members have in common.

TOPIC 7

Comparing, Ordering, and Equalizing on Numerousness

The children compare, order, and equalize sets on the attribute of numerousness. One-to-one matching is emphasized. Numbers and counting are not included.

REGULAR OBJECTIVES

1. Given two sets, compares them on the attribute of numerousness.
2. Given two sets, orders them on the attribute of numerousness.
3. Given two sets, equalizies them on the attribute of numerousness.
4. Given more than two sets, orders them on the attribute of numerousness.

TOPIC 8

Three Dimensional Shape

The children describe and classify three-dimensional objects on the attribute of shape.

REGULAR OBJECTIVE

1. Given a description of a shape, chooses a three-dimensional object that has that shape.

PREPARATORY OBJECTIVES

2. Given a three-dimensional object, describes its shape.
3. Given two containers, compares and orders them on capacity.

TOPIC 9

Representing Numerousness Physically

The children use objects to physically represent numerousness in situations in which direct comparing and ordering are impossible. They use various objects as representations and then compare, order, and equalize the representations to decide about the original sets.

REGULAR OBJECTIVES

1. Given a set, physically represents the numerousness of that set.
2. Given two or more sets and their physical representations, compares and orders those sets on the basis of the physical representations.

TOPIC 10

Paths and Location

The children work with locating objects. Paths are used as references.

REGULAR OBJECTIVES

1. Given a path and an object, describes the location of the object in relation to the path.
2. Given a path and a verbally identified location, places an object in that location.

PREPARATORY OBJECTIVES

3. Given a path, describes it.
4. Given several objects and an attribute, sorts the objects into sets on that attribute.

TOPIC 11

Representing Numerousness Pictorially

The children pictorially represent the numerousness of sets by graphing. Graphing serves as a transition between physically representing numerousness and symbolically representing numerousness.

REGULAR OBJECTIVES

1. Given a set, pictorially represents the numerousness of that set.
2. Given two or more sets and their pictorial representations, compares and orders the sets on the basis of the pictorial representations.

PREPARATORY OBJECTIVE

3. Given several objects and an attribute, sorts the objects into sets on that attribute.

TOPIC 12

Tallying

The children learn tallying which serves as a transition between pictorially representing numerousness (graphing) and symbolically representing numerousness (numbers).

REGULAR OBJECTIVE

1. Given a set of events or objects, represents the numerousness of the set with tally marks.

TOPIC 13

Time

The children compare and order events on time of duration and time of occurrence.

REGULAR OBJECTIVE

1. Given two events, compares and orders them on the attribute of time of occurrence.

PREPARATORY OBJECTIVES

2. Given an event, describes the time of its occurrence.
3. Given two or more events, compares and orders them on the attribute of duration.
4. Given a set, states an attribute that all of its members have in common.

TOPIC 14

Representing Numerousness Symbolically

Numerousness is represented symbolically by the number words and the number symbols. The children learn to count and to recognize the symbols. They do not learn to write the number symbols at this time.

REGULAR OBJECTIVES

1. Given a set of 1-10 objects, represents the numerousness of that set by stating the appropriate number.
2. Given a spoken or written number 0-10, represents it physically or pictorially.
3. Given a written numer 0-10, reads it.
4. Given a set of 0-10 objects and a set of written numerals, represents the numerousness of that set by choosing the appropriate written number.

PREPARATORY OBJECTIVE

5. Given a set of numbers 0-20, orders them.

TOPIC 15

Two Dimensional Shape

The children describe and classify regions including faces of solids on the attribute of shape.

REGULAR OBJECTIVES

1. Given a three-dimensional object, describes its shape.
2. Given an attribute, chooses a region that has that attribute.
3. Given two regions and an attribute, states whether the regions are the same or different on that attribute.

PREPARATORY OBJECTIVES

4. Given a region, describes its shape.
5. Given a region, represents it physically or pictorially.

TOPIC 16

Comparing and Ordering on Weight

The children compare and order real objects on the attribute of weight by using balances.

REGULAR OBJECTIVES

1. Given two objects, compares them on the attribute of weight.
2. Given two objects, orders them on the attribute of weight.

PREPARATORY OBJECTIVE

3. Given two objects, equalizes them on the attribute of weight.

TOPIC 17

Writing Numbers

The children learn to write the numbers 0-10 and practice writing them in a variety of situations.

REGULAR OBJECTIVE

1. Given a spoken number 0-10 or a set of objects or events, writes the number associated with the spoken number of set of objects or events.

PREPARATORY OBJECTIVES

2. Given a set of numbers 0-20, orders them.
3. Given an object or event in a sequence of objects or events, states the appropriate ordinal name for that object or event.

TOPIC 18

Comparing and Ordering Events on Time

The children compare and order events on time of duration and time of occurrence.

REGULAR OBJECTIVE

1. Given more than two events, orders them on the attribute of time of occurrence.

PREPARATORY OBJECTIVES

2. Given two or more events, compares and orders them on the attribute of duration.
3. Given two containers, compares and orders them on the attribute of capacity.
4. Given an object or event in a sequence of objects or events, states the appropriate ordinal name for that object or event.

TOPIC 19

Assigning Measurements

The children use arbitrary units to represent lengths or weights of objects by assigning a number and unit. Then they compare and order objects using these measurements.

REGULAR OBJECTIVES

1. Given an object, physically represents the weight of that object.
2. Given an object and a physical representation of a length of that object, assigns a measurement to that length.
3. Given an object and a physical representation of its weight, assigns a measurement to that weight.
4. Given the measurement of two lengths, compares and orders those lengths using the measurements.
5. Given the measurements of the weights of two objects, compares and orders those objects on weight using the measurements.

PREPARATORY OBJECTIVES

6. Given a set of numbers 0-10, orders them.
7. Given two sets, equalizes them on the attribute of numerousness.

TOPIC 20

Paths

The children describe a closed path (such as a triangle, rectangle, and square) in terms of number and length of sides. They are introduced to the use of the geoboard as a simple way to make paths.

REGULAR OBJECTIVES

1. Given a path, describes it.
2. Given a description of a path, chooses a path with that description.
3. Given a description of a path, constructs such a path.

PREPARATORY OBJECTIVE

4. Given two regions, compares and orders them on the attribute of area.

TOPIC 21

Comparison Sentences

The relationship between two sets or two objects (on a given attribute) is represented by a sentence involving = or ≠ (for example: $5 \neq 7$, $6 = 6$, $A \neq B$). The process of validating is introduced.

REGULAR OBJECTIVES

1. Given a comparison sentence, reads it.
2. Given two objects or sets, chooses an appropriate comparison sentence.
3. Given two objects or sets, writes an appropriate comparison sentence.
4. Given a comparison sentence, validates it physically or pictorially.
5. Given an open comparison sentence, completes it.

TOPIC 22

Comparing and Ordering on Capacity

The children directly compare and order the capacities of various containers by pouring from one to another. Also they learn to represent the capacity of a given container by assigning a measurement with arbitrary units.

REGULAR OBJECTIVES

1. Given two containers, compares and orders them on capacity.
2. Given a container, assigns a measurement to the capacity of that container.

PREPARATORY OBJECTIVES

3. Given the measurements of the capacities of two containers, compares and orders those containers using the measurements.
4. Given a set of numbers 0-20, orders them.

TOPIC 23

Order Sentences

The order relationship between two sets or two objects (on a given attribute) is examined further. Now if the two are not equal, the children decide which is larger and write an order sentence (for example: $5 < 7$, $6 = 6$, $A > B$). The children also learn to validate given order sentences.

REGULAR OBJECTIVES

1. Given an order sentence, reads it.
2. Given two objects or sets, chooses an appropriate order sentence.
3. Given two objects or sets, writes an appropriate order sentence.
4. Given an order sentence, validates it physically or pictorially.
5. Given an open order sentence, completes it.

PREPARATORY OBJECTIVE

6. Given a set of numbers 0-20, orders them.

TOPIC 24

The Numbers 0-20

The numbers 11-20 are introduced as representing the numerousness of sets of that many members. The children learn to recognize and write these numbers and to count such sets. The numbers 0-10 are reviewed.

REGULAR OBJECTIVES

1. Given a set of 11-20 objects or events, represents the numerousness of that set by stating the appropriate number.
2. Given a number 11-20, reads it.
3. Given a spoken number or a set of 11-20 objects or events, writes the number associated with that number or set of objects or events.
4. Given a number 11-20, represents it physically or pictorially.
5. Given a set of numbers 0-20, orders them.

TOPIC 25

Representing Equalizing Situations

The children learn to represent the process of equalizing by writing a sentence about how two sets or objects have been or will be equalized.

REGULAR OBJECTIVES

1. Given an open equalizing situation involving the numbers 0-20 in which the amount to be added on or taken away is unknown, writes a sentence which represents that situation.
2. Given an open equalizing situation involving the numbers 0-20 in which the amount to be added on or taken away is unknown, chooses a sentence which represents that situation.

PREPARATORY OBJECTIVE

3. Given a sentence involving the numbers 0-20, reads that sentence.

TOPIC 26

Movement and Direction

Simple maps are examined. The children follow simple oral or written directions involving movement on a given path or between given points. They also learn to give such instructions.

REGULAR OBJECTIVES

1. Given the names of points on a path in order, constructs such a path.
2. Given a path, describes the path by stating in order points on that path.

PREPARATORY OBJECTIVES

3. Given two objects, assigns a measurement to the distance between them.
4. Given an instruction to move in a specified direction, moves in that direction.
5. Given a figure, represents it physically.
6. Given a figure, represents it pictorially.

TOPIC 27

Representing Other Equalizing Situations

The children look at other equalizing situations in which one of the two sets to be equalized is unknown. They represent these situations with sentences such as $\square - 2 = 4$, $5 = \square + 3$, $2 + 6 = \square$, $\square = 10 - 3$. They also begin to solve sentences. The process of validating is stressed.

REGULAR OBJECTIVES

1. Given a sentence involving the numbers 0-20, reads it.
2. Given an open equalizing situation involving the numbers 0-20 in which one of the sets to be equalized in unknown, writes a sentence which represents the situation.
3. Given an open equalization situation involving the numbers 0-20 in which one of the sets to be equalized in unknown, chooses a sentence which represents that situation.

PREPARATORY OBJECTIVES

4. Given an open sentence involving the numbers 0-20, solves it.
5. Given a closed sentence involving the numbers 0-20, validates it physically or pictorially.

TOPIC 28

Symmetry, Fractions, and Shape

The children learn that some familiar figures are symmetric and told to test whether or not a given figure is symmetric. Fractions (halves, thirds, etc.) are introduced in terms of the attribute of area.

REGULAR OBJECTIVES

1. Given a figure, names it.
2. Given the name of a figure, chooses a figure with that name.

PREPARATORY OBJECTIVES

3. Given an object and a reference object, describes the location of the object in relation to that reference.
4. Given a figure, states whether it is symmetric with respect to some line.
5. Given an object or set divided into equal parts and given a designated subset of those parts, states the fractional name of the designated parts.
6. Given a fractional name and given an object or set divided into an appropriate number of fractional parts, represents that fractional name.
7. Given a set or object divided into parts and given a fractional name, states whether the set or object is divided into that fractional part.

TOPIC 29

Representing Joining and Separating Situations

The processes of joining and separating are introduced and represented by sentences. Active adding on and taking away are stressed. The children also solve such sentences.

REGULAR OBJECTIVES

1. Given an open joining or separating situation involving the numbers 0-20, writes a sentence that represents the situation.
2. Given a joining or separating situation involving the numbers 0-20, chooses a sentence that represents the situation.

PREPARATORY OBJECTIVES

3. Given an open sentence involving the numbers 0-20, solves it.
4. Given a closed sentence involving the numbers 0-20, validates it physically or pictorially.

TOPIC 30

Grouping

The children learn to separate any set by grouping the objects of the set into subsets of a given size. Then they record the results of their grouping. After grouping a set by 3, they might record 5(3) + 2.

REGULAR OBJECTIVES

1. Given a set of objects and an instruction to make groups of a specified size, appropriately groups the objects.
2. Given a grouped set of objects, writes the grouping notation that represents that set.

PREPARATORY OBJECTIVES

3. Given a grouping notation, represents it physically or pictorially.
4. Given a grouped set or its representation, and number to regroup by, appropriately regroups the objects.

TOPIC 31

Geometric Shapes

A variety of geometric ideas are reviewed and explored here with many types of materials.

PREPARATORY OBJECTIVES

1. Given a figure, represents it physically.
2. Given a figure, represents it pictorially.
3. Given an object or set divided into equal parts and given a designated subset of those parts, states the fractional name of the designated parts.
4. Given a fractional name and given an object or a set divided into an appropriate number of fractional parts, represents that fractional name.

TOPIC 32

Solving Number Sentences (0-10)

Other types of joining and separating situations are examined, and sentences are written to represent them. The children solve any given open sentence with numbers 0-10 and validate it. Mastery of number facts (0-10) is expected.

REGULAR OBJECTIVES

1. Given a difference situation involving the numbers 0-20, writes a sentence that represents that situation.
2. Given an open sentence involving the numbers 0-10, solves it.
3. Given a closed sentence involving the numbers 0-10, validates it physically or pictorially.

PREPARATORY OBJECTIVE

4. Given a part-whole situation involving the numbers 0-20, writes a sentence that represents that situation.

TOPIC 33

The Numbers 0-99

The numbers 21-99 are introduced as representing the numerousness of sets. The children learn to recognize and write these numbers in expanded [3(10) + 4] and compact (34) notation. Place value is introduced.

REGULAR OBJECTIVES

1. Given a set of 0-99 objects, represents the numerousness of that set by stating the appropriate number.
2. Given a set of 0-99 objects, represents the numerousness of that set by writing the appropriate number.
3. Given a number 0-99 written in expanded notation, writes it in compact notation.
4. Given a number 0-99, reads it.
5. Given a number 0-99, represents it physically or pictorially.
6. Given a number 0-99 written in compact notation, writes it in expanded notation.
7. Given a set of numbers 0-99, orders them.

TOPIC 34

Units of Capacity

Standard units of capacity (cup, quart, and gallon) are introduced. Capacities of various containers are represented, compared, and ordered in terms of these standard units.

REGULAR OBJECTIVE

1. Given a container, assigns a standard measurement to the capacity of that container.

PREPARATORY OBJECTIVES

2. Given measurements of the capacities of two containers, compares and orders those containers using those measurements.
3. Given an object or set divided into equal parts and given a designated subset of those parts, states the fractional names of the designated parts.
4. Given a fractional name and given an object or set divided into an appropriate number of fractional parts, represents that fractional name.
5. Given a set or object divided into parts and given a fractional name, states whether the set or object is divided into that fractional part.

TOPIC 35

Number Sentences (0-20)

Open sentences involving the numbers 11-20 are solved and validated. Mastery of number facts (11-20) is expected.

REGULAR OBJECTIVES

1. Given an open sentence involving the numbers 0-20, solves it.
2. Given a closed sentence involving the numbers 0-20, validates it physically or pictorially.

PREPARATORY OBJECTIVES

3. Given a part-whole situation involving the numbers 0-20, writes a sentence that represents that situation.
4. Given a set or object divided into parts and given a fractional name, states whether the set or object is divided into that fractional part.
5. Given a fractional name and given an object or set divided into an appropriate number of fractional parts, represents that fractional name.

TOPIC 36

Describing, Classifying, and Locating

The children sort objects on the basis of two or more attributes and are introduced to the intersection of sets. Finally they locate objects in relation to paths. They look at grids as intersections of columns and rows.

REGULAR OBJECTIVE

1. Given several objects and two or more attributes, sorts the objects into sets on the basis of at least two other attributes.

PREPARATORY OBJECTIVES

2. Given an object and a reference object, describes the location of the object in relation to that reference.
3. Given an object and a description of a location in terms of a reference object, places the object in that location.

TOPIC 37

Partitioning

The children divide a set into a given number of equal subsets and a remainder, and write the results in grouping notation. Then they apply this knowledge to fractions and divide various sets into halves, thirds, etc.

REGULAR OBJECTIVES

1. Given a set of objects and an instruction to partition that set into a given number of groups, appropriately partitions that set.
2. Given a partitioned set of objects, represents that set by writing grouping notation.
3. Given a grouping notation, represents it physically or pictorially.
4. Given an object or set divided into parts and a fractional name, states whether the object or set is divided into those fractional parts.
5. Given a fractional name and an object or set divided into an appropriate number of fractional parts, represents that fractional name.

PREPARATORY OBJECTIVE

6. Given two regions, compares and orders them on the attribute of area.

TOPIC 38

Number Sentences (0-99)

A variety of situations involving the numbers 0-99 are represented by sentences. The children solve open sentences about joining, separating, and differences using objects or pictures. The regrouping that occurs when the children are solving these problems is focused on as a background for algorithms (Topic 40).

REGULAR OBJECTIVE

1. Given an open equalizing, or joining, or separating situation involving the numbers 0-99, writes a sentence that represents that situation.

PREPARATORY OBJECTIVES

2. Given a grouped set of objects—or its representation—and an instruction to regroup, appropriately regroups the objects.
3. Given two numbers whose sum is 0-99, computes their sum.
4. Given two numbers 0-99, computes their difference.
5. Given an open sentence involving the numbers 0-99, solves it.

TOPIC 39

Units of Length

Standard units of length (inches, feet, yards, centimeters, and meters) are introduced. The children measure various lengths using these units and compare and order the measurements.

REGULAR OBJECTIVE

1. Given an object, assigns a standard measurement to a length of that object.

PREPARATORY OBJECTIVES

2. Given standard measurements of two lengths, compares and orders those lengths using the measurements.
3. Given a common fraction and an object or set divided into an appropriate number of fractional parts, represents that fraction by using that object or set.
4. Given an object or set divided into equal parts and a designated subset of those parts, chooses or writes an appropriate common fraction that represents the designated parts.
5. Given a common fraction, reads it.

TOPIC 40

The Addition and Subtraction Algorithms (0-99)

Addition and subtraction of two-digit numbers with and without regrouping are introduced. The children use their background with regrouping objects in Topic 38 to give meaning to the algorithms.

REGULAR OBJECTIVES

1. Given two numbers whose sum is 0-99, computes their sum.
2. Given two numbers 0-99, computes their difference.

PREPARATORY OBJECTIVE

3. Given an open addition or subtraction sentence involving the numbers 0-99, states how to solve it.

National Evaluation Committee

Francis S. Chase, Chairman
Emeritus Professor
University of Chicago

Helen Bain
Past President
National Education Association

Lyle Bourne
Professor
University of Colorado

Sue Buel
Consultant, Portland, Oregon

Roald F. Campbell
Emeritus Professor
The Ohio State University

George E. Dickson
Dean, College of Education
University of Toledo

Larry R. Goulet
Professor
University of Illinois

Chester W. Harris
Professor
University of California - Santa Barbara

William G. Katzenmeyer
Professor
Duke University

Barbara Thompson
Superintendent of Public Instruction
State of Wisconsin

Joanna Williams
Professor
Teachers College
Columbia University

University Advisory Committee

John R. Palmer, Chairman
Dean
School of Education

William R. Bush
Deputy Director
R & D Center

David E. Cronon
Dean
College of Letters and Science

Diane H. Eich
Specialist
R & D Center

Evelyn L. Hoekenga
Coordinator
R & D Center

Dale D. Johnson
Associate Professor
Curriculum and Instruction

Herbert J. Klausmeier
Member of the Associated Faculty
R & D Center

James M. Lipham
Member of the Associated Faculty
R & D Center

Wayne R. Otto
Associate Director
R & D Center

Richard A. Rossmiller
Director
R & D Center

Elizabeth J. Simpson
Dean
School of Family Resources
and Consumer Sciences

Len Van Ess
Associate Vice Chancellor
University of Wisconsin - Madison

Associated Faculty

Vernon L. Allen
Professor
Psychology

B. Dean Bowles
Professor
Educational Administration

Thomas P. Carpenter
Assistant Professor
Curriculum and Instruction

Marvin J. Fruth
Professor
Educational Administration

John G. Harvey
Professor
Mathematics
Curriculum and Instruction

Frank H. Hooper
Professor
Child Development

Herbert J. Klausmeier
V. A. C. Henmon Professor
Educational Psychology

Joseph T. Lawton
Assistant Professor
Educational Psychology

Joel R. Levin
Professor
Educational Psychology

L. Joseph Lins
Professor
Institutional Studies

James M. Lipham
Professor
Educational Administration

Donald N. McIsaac
Professor
Educational Administration

Gerald Nadler
Professor
Industrial Engineering

Wayne R. Otto
Professor
Curriculum and Instruction

Robert G. Petzold
Professor
Music
Curriculum and Instruction

Thomas S. Popkewitz
Assistant Professor
Curriculum and Instruction

Thomas A. Romberg
Professor
Curriculum and Instruction

Richard A. Rossmiller
Professor
Educational Administration

Dennis W. Spuck
Assistant Professor
Educational Administration

Michael J. Subkoviak
Assistant Professor
Educational Psychology

Richard L. Venezky
Professor
Computer Sciences

J. Fred Weaver
Professor
Curriculum and Instruction

Larry M. Wilder
Assistant Professor
Child Development